THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XVI | April, 1925 | Number 4

THE USE OF TESTS IN THE EVALUATION OF FACTORS WHICH CONDITION THE ACHIEVEMENT OF PUPILS IN HIGH SCHOOL PHYSICS

J. M. HUGHES

Northwestern University

I

Intelligence as a Factor Conditioning Achievement

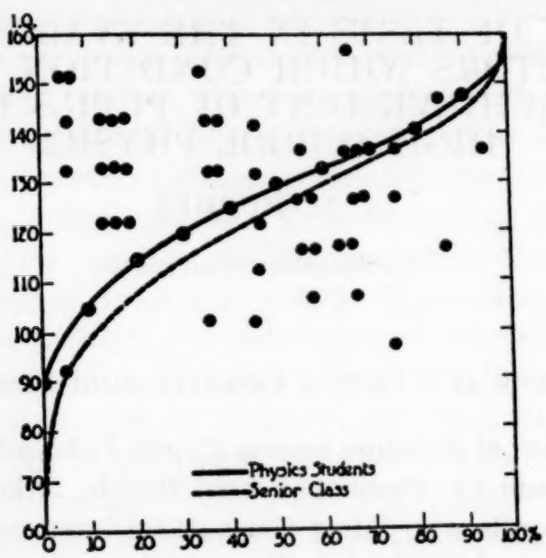
The Conception of Selection among Pupils Taking Physics.—Studies such as that made by Powers at Fort Smith, Arkansas,[1] show the intelligence of pupils who select the natural sciences to be relatively high when compared to the intelligence of pupils who select certain other subjects. Common opinion has, likewise, held this to be true. It is more likely, if one studies schools singly, that he will find an evidence of some mental selection on the part of the pupils taking physics. A comparison of the intelligence of the pupils taking physics in one of the high schools of Minneapolis with the intelligence of the whole of the class from which the pupils in physics came is shown in Fig. 1. The intelligence of the pupils taking physics is seen to be distinctly superior to that of the whole group. If a teacher should generalize on the basis of such a study or a few such studies, he would feel justified in having his pupils busy themselves with the more abstract, theoretical, and logical phases of the subject. Such material, he may contend, would be of most profit to them since it is most nearly consonant with their abilities.

Selection in Different Schools May Not Be the Same.—Selection may not be positive in every school as the few studies made suggest; there are too many factors determining selection to cause it to be uniform

[1] Powers, S. R.: Intelligence as a Factor in the Selection of High School Subjects. *School Review*, Vol. XXX, pp. 452–455, June, 1922.

in nature in every school. The policy of a teacher in grading, his skill and reputation in making the subject interesting, the place the subject occupies in the curriculum, and other conditions may determine the nature of the selection, and cause it to be positive, negative, or normal.

The Degree of Selection of the Two Sexes May Not Be the Same.— There may be selection within a selection in the group. The girls, on the average may not be of equal intelligence to the boys. They are



Distribution of 45 High School Pupils in Physics and entire Senior Class in a Minneapolis High School from which selection was made

FIG. 1.

apparently inferior to boys in their achievement in physics. This could be due to a lower average intelligence.

Some Data upon the Degree of Selection of Pupils Taking Physics.— The intelligence quotients of 1036 pupils taught by 30 teachers in 29 schools were procured. A distribution of the schools on the basis of size may be seen in Table I.

TABLE I.—DISTRIBUTION OF THE SCHOOLS ON THE BASIS OF SIZE FROM WHICH INTELLIGENCE DATA WERE PROCURED

ENROLLMENT OF SCHOOL	NUMBER PARTICIPATING
100–249	6
250–399	5
400–549	3
550–999	4
1000–1899	5
1900–3500	6
Total	29

The Otis Self-administering Test of Mental Ability, Higher Examination: Form A, was used as the basis for the study of the intelligence factor. This test was chosen because, in addition to possessing all of the attributes demanded of a reliable standard test, it may be satisfactorily administered by teachers who are more or less untrained in the technique of administering mental tests. Most of the intelligence quotients were obtained by the Otis Test, having been read from the chart accompanying the manual for these tests.

The Translation of Test Data into Otis Equivalents.—Data on tests other than the Otis were procured from five schools. These were

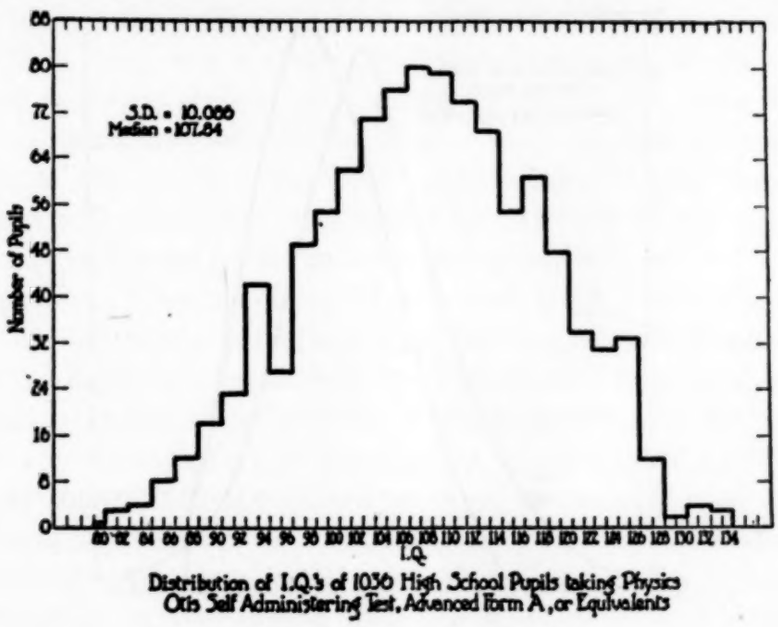

S.D. = 10.066
Median = 107.84
Number of Pupils
I.Q.
Distribution of I.Q.'s of 1036 High School Pupils taking Physics
Otis Self Administering Test, Advanced Form A, or Equivalents

FIG. 2.

translated into IQs having the same central tendency and dispersion as the IQs on the Otis Self-administering Test. Such a procedure is necessary for direct comparison of IQs since neither the central tendency nor the variability of any two group tests is the same. This translation was facilitated by means of a graph described in an article by Dr. W. S. Miller, Professor of Educational Psychology, the University of Minnesota.[1]

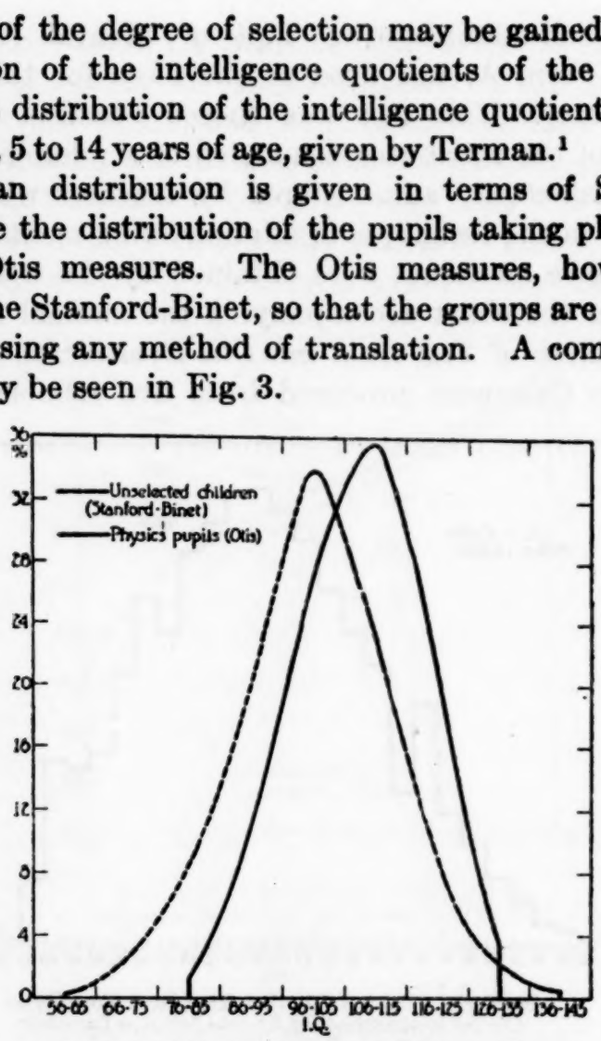
The Degree of Selection.—The intelligence quotients for 1036 pupils were calculated. The total distribution is shown in Fig. 2.

---

[1] Miller, W. S.: Variation and Significance of IQs Obtained from Group Tests of Intelligence. *Journal of Educational Psychology*, September, 1924.

Some idea of the degree of selection may be gained by comparing the distribution of the intelligence quotients of the pupils taking physics with a distribution of the intelligence quotients of 905 unselected children, 5 to 14 years of age, given by Terman.[1]

The Terman distribution is given in terms of Stanford-Binet measures while the distribution of the pupils taking physics is given in terms of Otis measures. The Otis measures, however, closely approximate the Stanford-Binet, so that the groups are fairly comparable without using any method of translation. A comparison of the two groups may be seen in Fig. 3.



Distribution of I.Q.'s of 905 unselected Children, 5-14 Years; Stanford-Binet Tests (Terman) and of 1036 Pupils taking Physics; Otis Self Administering Tests; Advanced Examination, Form A

FIG. 3.

The distribution for the pupils taking physics shows a greater percentage of the pupils to be above normal intelligence than is the case in the distribution of unselected children. This does not seem to be true at the upper end of the curve. This can possibly be explained as being due to some characteristic of the Otis Test. It seemingly is not constructed to measure the intelligence of older pupils who possess very high intelligence.

---

[1] Terman, Lewis M.: "The Measurement of Intelligence." p. 66.

Pupils in the physics group who have intelligence quotients of 96 or above comprise 88 per cent of the group as compared with 68.67 per cent of the unselected children. Twenty-four and six-tenths per cent of the pupils taking physics have the superior intelligence quotients of 116 or above while but 11.85 per cent possess this degree of intelligence in the unselected group. While this selection is significant, still the fact remains that there is a large amount of overlapping in the distributions of the two groups. There is not a great deal of difference between the ranges of the two groups. One-eighth of the pupils taking physics had Stanford-Binet Intelligence Quotients of 95 or below, and no insignificant per cent had intelligence quotients of less than 90. If physics material and instruction are being adapted to pupils of superior ability only, certainly the subject can easily be made too difficult and too abstract for a very significant proportion of the pupils taking it. However, approximately two pupils who take physics will come from the upper 50 per cent of the population where but one will come from the lower 50 per cent.

Comparison of Pupils Taking Physics with High School Juniors and Seniors.—We are more concerned with the degree of selection of the pupils taking physics from among the juniors and seniors in the high schools than with the selection from the population at large. A comparison of the distribution of intelligence quotients of pupils taking physics was made with the distribution of the intelligence quotients of the juniors and seniors in six different high schools. Data for the latter were taken from a study made by Madsen[1] on the selective factor operating in the junior and senior classes of different high schools. Madsen's study included 2608 juniors and seniors in high schools in six different cities. His data give the median scores these pupils made on the Army Alpha Test. By taking the median scores of each group given in Madsen's data and by taking into account the number of pupils included in each group, we are able to find the mean score for junior and senior boys, for junior and senior girls, and for all juniors and seniors. By using the formula $.8S + 108$ for finding mental age in months from the raw score on the Army Test, the mean mental age of each group can be computed. Since a number of studies[2]

[1] Madsen, Iver N.: Group Intelligence Tests as a Means of Prognosis in High School. *Journal of Educational Research*, Vol. III, pp. 43–52.

[2] This is substantiated in an unpublished age grade study made by Fred von Borgersrode, Assistant Director of the Bureau of Educational Research of the University of Minnesota of 103 high schools in Minnesota made in 1923–1924 including approximately 12,000 pupils.

show the mean chronological age of juniors and seniors to be above 16, and since 16 was used as a basis for the norms of the Army Alpha Test, we may take 16 as the chronological age of juniors and seniors. We may then compute the mean IQs for the different groups. By translating the IQs to the Otis equivalents we have comparable groups. The data for comparison are given in Table II.

TABLE II.—COMPARISON OF THE INTELLIGENCE LEVELS OF JUNIORS AND SENIORS OF SIX CITIES IN MADSEN'S STUDY WITH THE INTELLIGENCE LEVELS OF JUNIORS AND SENIORS TAKING PHYSICS

	Number	Mean IQ on Alpha	Mean IQ on Otis
Madsen's data			
Junior and senior boys	1065	112.6	109
Junior and senior girls	1543	106.5	105
Mean for both	2608	108.1	106.1
Data in this study			
Junior and senior boys	771	.....	108.18
Junior and senior girls	265	.....	107.55
Mean for boys and girls	1036	.....	108

The mean IQ of the boys taking physics is less than one point below the mean IQ of the boys included in Madsen's study. The difference is too small to be significant.

The mean IQ of the girls taking physics is 2.55 points above the mean IQ of the girls included in Madsen's study. In considering this difference, we must not fail to recognize the slight inappropriateness of using the Alpha Tests as a basis of comparison for the girls. It is commonly believed that these tests underrate the ability of girls.

In order to add to the reliability of the findings in our comparison with the Madsen study, a similar comparison was made with data furnished by Miller's study of the intelligence of pupils in different grades in a number of high schools in Minnesota.[1] The data as obtained by Miller are given in Table III.

[1] The data here were furnished the writer by Dr. W. S. Miller, Professor of Educational Psychology, the University of Minnesota.

TABLE III.—THE INTELLIGENCE LEVELS OF BOYS AND GIRLS IN JUNIOR AND SENIOR CLASSES IN MINNESOTA HIGH SCHOOLS (AFTER MILLER)

Boys	
Number of boys in junior year	306
Mean score of junior boys on Miller A	67.8
Number of boys in senior year of high school	188
Mean score of senior boys on Miller A	70.9
Mean score of boys in junior and senior years of high school	69
Mean mental age of boys by formula, $2S \div 80$	218 months
Mean IQ of boys in junior and senior years of high school	Miller 113.5
Mean IQ of boys in junior and senior years of high school	Otis 108.2
Girls	
Number of girls in junior year	467
Mean score of junior girls on Miller A	68.9
Number of girls in senior year of high school	198
Mean score of senior girls on Miller A	72.6
Mean score of girls in junior and senior years of high school	70.7
Mean IQ of girls in junior and senior years of high school	Miller 115.3
Mean IQ of girls in junior and senior years of high school	Otis 109

The comparisons point to the conclusion that the boys who take physics are not a more highly selected group than are the pupils of the junior and senior classes. Since boys comprise more than 70 per cent of the pupils taking physics, this conclusion holds at least for the major portion of the group.

The Miller and Madsen studies do not agree as to the degree of selection among the girls. The mean IQ of the girls in Madsen's study was equivalent to an Otis IQ of 105. The mean IQ of the girl's in Miller's study was equivalent to an Otis IQ of 109. The mean IQ of the girls involved in this study was 107.5 on Otis. There is little likelihood that girls who take the subject of physics in high school are superior in ability to those who are eligible to elect the subject but do not choose to do so.

The data as a whole seem to indicate that, taken as a group, the pupils in the physics classes are fairly representative of the enrollments of the junior and senior classes as a whole. There is the same degree of variability, the same wide range of abilities, and the same frequency of grades of ability above and below the average.

While the total group does not show any selection, this should not be interpreted to mean that individual schools would not show some kind of selection in their classes in physics. Evidently, as in other subjects, as has been mentioned, the policy of grading, likes or dislikes for the teacher, the place of the subject in the curriculum, and

similar factors, would somewhat affect the type of pupil choosing the subject. Even within the same school the classes of different teachers show different degrees of selection.

Intelligence data were received from three schools in each of which two teachers of physics were employed. In none of the schools was the average intelligence of the pupils of the different teachers the same. In each there were differences of from five to seven points.

Appropriateness of Material to the Intelligence of the Pupil.—Contention is often made that the pupil will probably profit most from a subject when the content of the subject is consonant with his ability. There are certain materials in physics which are probably appropriate to pupils of every grade of intelligence found in the upper years of the high school. There are certain other materials which are probably inappropriate to pupils with the lower grades of intelligence. It has been shown that there are many pupils taking physics who have intelligence quotients less than 100. There are likewise many pupils with very superior intelligence quotients of 115 or above. Certainly, there are materials appropriate to pupils of both groups. On the other hand, there may be materials which may be readily comprehended by those of superior intelligence and comprehended but little by those of lower intelligence. It is not without the realm of possibility that some of the content of the physics course is difficult of comprehension by pupils of even the highest intelligence.

The availability of tests of intelligence makes it possible to measure the extent to which intelligence enters into the comprehension of specific content.

Comparison of Achievement of Two Groups of Pupils of Different Average Intelligence.—The pupils were classified into two groups. One group included only those pupils with intelligence quotients of 116 and above. The other group included only those pupils with intelligence quotients of 99 and below. One hundred pupils, chosen at random, were included in each group. Two such groups were chosen for each of three tests which cover three divisions of the subject of physics.

In the mechanics and heat group, 121.5 and 92.72 were the averages of intelligence quotients of the pupils in the respective groups. In the magnetism and electricity groups the averages were 121.72 and 94.22, respectively. In the sound and light groups, the averages were 120.4 and 94.42, respectively. The two groups in each test differ in average intelligence by a little more than 25 points.

The two groups differing, then, in the average in intelligence quotients by about 25 points, were compared as to the number of correct responses made by the pupils in each group to 25 questions selected from the writer's tests in physics. Eight questions were taken from a test in mechanics and heat, eight from a test in magnetism and electricity, and four from a test in sound and light. The questions as stated in the tests follow. They are stated in the order of difficulty.

1. In the formula D *equals* $\frac{M}{V}$, M means mass, V means volume. What does D mean?

2. What name is given the device on the engine of an automobile which changes the liquid fuel into vapor and mixes air with it in proper proportion for complete combustion?

3. In the following sentence, draw a line under the one of the six words at the end that makes the truest sense.

The amount of heat required to raise the temperature of 1 gram of water 1° Centigrade is called mechanical, equivalent, specific heat, heat of fusion, calorie, latent heat.

4. One foot of a certain wire has a resistance of 1 ohm. What will be the resistance of 10 feet of the same wire?

5. Underline in the following a good conductor of electricity: Wool, glass, dry air, paraffin, silver, sulphur, resin.

6. What name is given to the bending of the light wave as it goes from one substance into another of different density?

7. What kind of cells are used on automobiles for starting, lighting, and ignition purposes?

8. A reflected sound wave returns to its source in .5 of a second. How many times as far away was a second reflector from which the sound returned in 1 second?

9. A current of 4 amperes is flowing through a variable resistance. The resistance is multiplied four times. What is the current?

10. What method of heat transference is utilized in the method of heating buildings with a hot air furnace?

11. What do we call the time rate of doing work?

12. What is the resistance of a 110-volt incandescent lamp, it if requires a current of .5 ampere to light it?

13. Two forces of 30 and 40 pounds, respectively, act at right angles. What is the resultant?

14. An observer moves toward a source of sound until he reaches a point one-third his original distance from it. As the sound now reaches the ear of the observer, how many times does its intensity appear to be multiplied?

15. A tall standpipe in Minnesota at noon time on a sunshiny day in summer bends a few hundredths of an inch. In what direction does it bend?

16. How many times will the heating effect of a wire be multiplied if you triple the current running through it?

17. How many foot pounds of work can be done by a two horse-power engine in eight hours?

18. A 16-candle power light is placed 9 feet from a screen. How many feet from the screen must one place a 4-candle power light in order to give the same intensity of illumination as is obtained from the 16-candle power light?

19. How much current will six cells connected in combination of three in series and two in multiple, send through a resistance of 2.91 ohms, if the electromotive force of each cell is 1 volt and the internal resistance of each cell is .06 ohm?

20. If the resistance between two points is 25 ohms, what will be the resistance of an additional wire connecting these points, in order to make the joint resistance between the points only 20 ohms?

A comparison of the responses of the two groups of pupils may be seen in Table IV.

The PE values and percentages of correct responses on the 20 questions are given for approximately 1400 unselected pupils who had taken the tests. The same data are given for the 100 selected pupils with IQs of 115 and above and for the 100 selected pupils with IQs of 99 or below.

In every question the percentage of pupils giving correct responses was less in the group with low IQs than in the group with high IQs. The average difference in the percentage of correct responses between the two groups is 21.5 per cent. When material of the difficulty represented by questions 18, 19, and 20 is reached, the pupils of lower intelligence compare very unfavorably with those of higher intelligence. Six pupils of superior intelligence correctly solve problem 18 to where one of lower intelligence correctly solves it. The ratio in problem 19 is almost fourteen to one. In problem 20 it is six to one. Equal or greater ratios are found for all of the questions of a more difficult nature in the tests.

TABLE IV.—THE PE VALUES AND PERCENTILE VALUES OF 20 QUESTIONS BASED UPON THE CORRECT RESPONSES OF APPROXIMATELY 1400 UNSELECTED PUPILS, 100 SELECTED PUPILS WITH IQS OF 115 AND ABOVE, AND 100 SELECTED PUPILS WITH IQS OF 99 OR BELOW

Number of question	PE values based upon 1400 pupils	PE values based upon 100 pupils with IQ of 115 or above	PE values based upon 100 pupils with IQ of 99 or below	Percentage of 1400 unselected pupils who gave correct responses	Percentage of 100 pupils with IQ of 115 or above who gave correct responses	Percentage of 100 pupils with IQ of 99 or below who gave correct responses	Difference in percentage of correct responses
1	+1.997	+1.475	+1.248	91.1	84	80	4
2	+1.518	+2.188	+ .735	84.7	93	69	24
3	+1.481	+1.475	+ .909	84.1	84	73	11
4	+1.456	+1.742	+1.248	83.7	88	80	8
5	+1.409	+1.900	+1.357	82.9	90	82	8
6	+1.096	+1.145	+ .571	77.0	78	65	13
7	+1.052	+2.083	+ .820	76.1	92	71	21
8	+ .364	+ .693	− .149	59.7	68	46	22
9	+ .160	+1.145	− .112	54.3	78	47	31
10	− .030	+ .778	− .531	49.2	70	36	34
11	− .067	+ .187	− .337	48.2	55	41	14
12	− .153	+1.000	− .612	46.9	75	34	41
13	− .160	− .453	−1.602	45.7	38	14	24
14	− .213	− .047	−1.096	44.3	48	23	25
15	− .519	− .261	− .376	36.3	57	40	17
16	− .660	− .037	− .778	32.8	49	30	19
17	− .842	+ .074	−1.302	28.5	52	19	33
18	−1.140	− .778	−2.439	22.1	30	5	25
19	−1.677	− .909	−3.044	12.9	27	2	25
20	−2.562	− .492	−2.305	4.2	37	6	31

The Need of Information as to the Intelligence of Classes and of Individuals.—The questions chosen for comparison would not be judged by teachers of physics to be difficult. Yet every question, after the seventh in the list, was answered by less than 50 per cent of the pupils in the lower intelligence group. A glance at the first seven questions will show that they are very simple and involve very simple thought processes. Some would contend that these questions are merely every day fact questions and scarcely worthy of being termed "questions in physics."

On the other hand 16 of the first 17 questions were answered by 48 per cent or more or the high intelligence group. Surely, under proper instruction, this group could have done much better. It may be well-

nigh impossible to get such accomplishment from those in the lower intelligence group.

The intelligence of the pupils conditions their achievement in physics. Apparently much of the material in the present course is too difficult for pupils of the lower degrees of intelligence. Under present instruction, much of the more difficult material is not being mastered by pupils of higher degrees of intelligence. For economy of time, it seems that the more abstract phases of the high school course in physics may well be omitted for most pupils. Other materials, applicable but to those of higher intelligence, should be selected by the teacher, only after he is satisfied that the material chosen is appropriate. To do this intelligently, it seems both necessary and obligatory that the teacher know the intelligence of his pupils individually and as a group. Thus the teacher whose pupils were found in this study to have a mean IQ of 116.5 had a different teaching situation from the teacher whose class was found to have a mean IQ of 96.2. The teacher of the latter group lamented the fact that his pupils did not do as well on the tests as he thought they should, stating at the same time that it was his intention at least to see that they completed all of the work of the text. Neither the materials nor the methods of these two classes should be the same. Yet both of the teachers were teaching without any measures of intelligence of their pupils and both were expecting a degree of achievement that would be superior to that ordinarily found in the class room. The first was succeeding admirably and the second was failing miserably.

It may well be supposed that a more select class may have pupils in it who do not possess high intelligence quotients. Injustice is likely to be done to these pupils when results are demanded which, for them, are impossible.

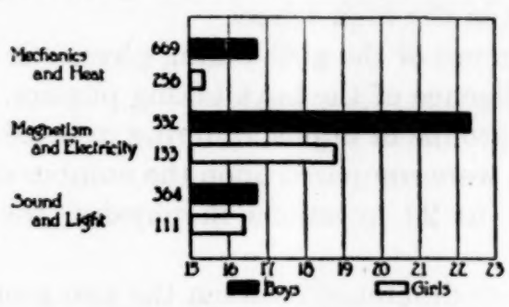
Intelligence, then, conditions the achievement of the pupil in physics. It should condition the choice of materials and the methods to be employed.

Appropriateness of the Material to the Sex of the Pupil.—The method used in determining the appropriateness of material to the intelligence of the pupil by taking specific items may well be extended to determining the appropriateness of material to the sex of the pupil. Before doing this, however, we should determine whether the material as taught does appeal more to one of the sexes than to the other. Results of different studies have shown a considerable difference in achievement between the boys and the girls. Camp, in his manual

for the Iowa Physics Scales, gives norms for boys and norms for girls. The norms for the boys are considerably higher than those for the girls. A similar difference between the achievement of boys and of girls was found in the tests used in this study. The extent of this difference may be seen from a study of Fig. 4.

The boys show superior achievement in all three tests. The difference between the achievement of the two sexes on the test in mechanics and heat was 1.4 points or .59Q. In magnetism and electricity the boys were superior by 3.5 points or .94Q. In sound and light the boys were superior by .3 point. This last difference is insignificant.

The intelligence quotients of the boys and girls represented in the data given above were about equal. Therefore the difference in



Differences in Achievement between Boys and Girls of approximately the same Intelligence in the three Divisions of Physics

FIG. 4.

achievement is not due to differences in intelligence. It seems that, if the subject of physics is not going to be adapted to the needs of the girls who take it, no school should make the subject a required subject for them. Certainly, where classes consist of a fair proportion of girls, efforts should be made to adapt the materials of instruction to their needs and interests.

The teacher then may not only adapt his methods and materials of instruction to the intelligence of his pupils, but he may go further, and adapt them to the needs and interests of the pupils of both sexes as they are represented within the group.

SUMMARY

1. Overestimation of the intelligence of pupils taking physics is likely to lead the teacher to emphasize the more abstract and theoretical phases of the subject.

2. The degree of selection of pupils in one school may not be a criterion for judging the degree of selection in other schools. The policy of the teacher in grading, his skill in making the subject interesting, its place in the curriculum, may affect the degree of mental selection among the pupils taking it.

3. When compared to the population at large, the pupils taking physics are a select group. Eighty-eight per cent of the group taking physics as represented in this study had intelligence quotients of 96 or above as compared with 68.7 per cent of unselected children as given by Terman.

4. There is no evidence in this study that there is any selection among the pupils taking physics from among the pupils in the junior and senior classes in the high school.

5. The intelligence of the girls taking physics is approximately the same as the intelligence of the boys taking physics.

6. When two groups of pupils, differing in intelligence by approximately 25 points, were compared upon the number of correct responses each group gave to 20 questions in physics, the following results were apparent.

(*a*) The average difference between the two groups in percentages of correct responses was 21.5.

(*b*) The group of pupils of lower intelligence gave a smaller percentage of correct responses to each of the 20 questions than the group of higher intelligence.

(*c*) The group with lower intelligence compared very unfavorably with those of higher intelligence in the percentages of correct responses given to questions of greatest difficulty.

(*d*) Excepting the first seven questions, which were very simple, no question was answered by as many as 50 per cent of the group of lower intelligence, while 16 of the first 17 questions in the list were answered by 48 per cent or more of the group of higher intelligence.

7. In order to adapt properly the materials and methods of the course to the intelligence of the pupils, it seems both necessary and obligatory that the teacher know the intelligence of his pupils individually and as a group.

8. The achievement of girls as measured by tests used in this study is inferior to the achievement of boys in physics in mechanics and heat and in electricity and magnetism. There was an insignificant difference in the achievement of the sexes in sound and light. Since the intelligence of the two sexes was approximately the same, the difference

in achievement seems to be explained through a difference in interests of the two sexes.

9. As physics is now taught, a requirement of the subject of girls as made in many schools seems unjustifiable.

10. There is a need for effort on the part of teachers of physics to adapt their materials and their instruction to the intelligence of their pupils and to adapt these more equitably to the needs and interests of both sexes.

THE RELIABILITY COEFFICIENT OF PERSONAL RATINGS[1]

EUGENE SHEN

Stanford University

The reliability of mental tests is usually measured by a correlation between results of two comparable tests. By analogy, the reliability of personal ratings may be evaluated by a correlation between ratings by two comparable judges. By pairing elements of two tests such that they are similar in difficulty and type, an author can to a certain extent insure the comparability between his tests. But the comparability of judges is much more precarious; it is entirely beyond the control of the investigator except by a meager selective function that he may fallibly exercise. On account of this uncontrollable variability of judges, a correlation between two judges is a very crude approximation to the reliability of either. The reliability of a judge thus crudely evaluated often varies considerably according to the judge with whom he happens to be correlated.

In a study on personal ratings, 28 individuals were requested to rank themselves and one another with respect to eight different traits. For reasons which do not concern us here, the 28 men were divided into four groups, each rating a combination of only four of the eight traits. And because two persons failed to turn in their ratings, our final data consisted of 13 series of ranks for each of the eight traits. These ranks were then converted into scores in terms of standard deviations of a unit normal distribution, and ratings by judges of the same group on the same trait were correlated with one another, yielding 228 coefficients in all. In spite of the favorable condition that all the individuals had been classmates for at least three years, the correlations showed an extreme variability. Those of ratings on memory, for instance, extended over a range of from .84 to −.04. The difficulty of securing comparable judges is evident.

Since the judges are far from comparable, the evaluation of the reliability of a single judge becomes a problem. As an extreme instance, the correlations of Judge V's ratings on *impulsiveness* with those by other judges vary from .70 to −.09. What should one say is his reliability? The most simple and natural answer is to take his

[1] Slightly modified from parts of a thesis written under the direction of Prof. Truman L. Kelley, to whom grateful acknowledgment is due for helpful suggestions and constant encouragement.

average correlation with the other judges. True, the average is closer to the truth than any one individual correlation coefficient. But it need not be the best approximation.

When a series of ratings with a true reliability 1.00 is correlated with another having a true reliability .01, the coefficient obtained is .10. The obtained coefficient is always the result of a fusion of, or a pulling between, the true reliability coefficients. Except when the two series are of truly equal reliability, the obtained coefficient is constricted for one and dilated for the other. Averaging does not eliminate this effect of pulling unless the different judges with whom the judge in question is correlated distribute themselves evenly in respect to reliability around him. Since it is impossible for judges of varying reliability to distribute themselves reciprocally around one another, the average correlation of one judge with the others as an approximation to his reliability retains a systematic error. Our task now is to derive a formula for the elimination of the systematic error, which we may call "correction for pulling."

The fundamental basis for our correction for pulling is the following theorem: *The correlation between two series of ratings on the same trait, independent in errors of each other, is equal to the geometric mean of their true reliabilities.* This can easily be demonstrated by means of the Spearmen correction formula for attenuation,

$$r_{\infty\omega} = \frac{r_{12}}{\sqrt{r_{11}}\sqrt{r_{22}}}$$

Since both series are on the same trait, $r_{\infty\omega}$ of necessity equals unity. Hence,

$$\frac{r_{12}}{\sqrt{r_{11}}\sqrt{r_{22}}} = 1 \quad \text{or} \quad r_{12} = \sqrt{r_{11}}\sqrt{r_{22}}$$

Similarly, $r_{13} = \sqrt{r_{11}}\sqrt{r_{33}}$ and $r_{23} = \sqrt{r_{22}}\sqrt{r_{33}}$

Multiplying the first two equations and dividing by the third, we have

$$r_{11} = \frac{r_{12}r_{13}}{r_{23}}$$

or more generally,

$$r_{11} = \frac{r_{1p}r_{1q}}{r_{pq}}$$

This formula is applicable whenever three independent series of ratings are obtained. It should be noted, however, that a reliability coefficient thus derived may have values outside the limits +1 and −1. In fact, when r_{pq} happens to be zero, the value for r_{11} will be infinite. Apparently a single determination from three obtained coefficients may give a reliability coefficient that is itself very unreli-

able. If, as in our present instance, we have six or seven judges in a group, there are 10 or 15 such determinations possible, and some method of a composite determination is expected to give more reliable results. But here again, a simple average will not do, as infinite values cannot be thus eliminated. The theoretically best method conceivable is to obtain an average from individual determinations weighted inversely as the squares of their standard errors. The standard error of a reliability coefficient corrected for pulling is derived by taking logarithmic differentials, squaring, summing, dividing by N, and substituting values as given by formulas (108*b*) and (129) in Kelley's "Statistical Method." The final equation after simplification is here given:

$$\sigma_{r_{1I}} = \frac{r_{1I}}{\sqrt{N}}\left[4r_{1I} + \frac{2}{r_{1I}} + \frac{1}{r^2_{1p}} + \frac{1}{r^2_{1q}} + \frac{1}{r^2_{pq}} - \frac{2r_{12}}{r^2_{1p}} - \frac{2r_{12}}{r^2_{1q}} - 5\right]^{\frac{1}{2}}$$

As the foregoing equation shows, any value for r_1 that is infinite has an infinite standard error and will accordingly be given a zero weight. Likewise other unreasonable values with large standard errors will receive small weights. The weighted average will thus have the minimum standard error and is undoubtedly the best value for a reliability coefficient.

As a practical procedure, however, the labor involved is enormous. To determine the reliability of seven judges would require the calculation of 21 coefficients of correlation, 105 corrections for pulling, 105 standard errors, and 105 weightings.

In view of the laborious procedure necessary for the calculation of the weighted average, we shall consider a simpler but hardly less accurate evaluation. We begin from the same fundamental equation,

$$r_{12} = \sqrt{r_{1I}}\sqrt{r_{2II}}$$

But instead of using all correlations of one judge with each of the others, we shall treat the correlation between one judge and the composite of the others as a whole. Letting the subscript $(S-1)$ denote the composite of all the judges except the one in question, the equation becomes

$$r_{1(S-1)} = \sqrt{r_{1I}}\sqrt{r_{(S-1)(S-1)}} \qquad \text{or} \qquad r_{1I} = \frac{r^2_{1(S-1)}}{r_{(S-1)(S-1)}}$$

Substituting values for the numerator and denominator as given by formulas (175) and (157), respectively, in "Statistical Method," and simplifying, we have

$$r_{1I} = \frac{(a-2)\bar{r}^2_{1P}}{a\bar{r}_{Pq} - 2\bar{r}_{1P}}$$

where a is the number of judges, $\bar{r}_{1P}$ the average correlation of the judge in question with the others, and $\bar{r}_{Pq}$ the average intercorrelation of all the judges.

It is satisfying to see that this formula will under all ordinary circumstances yield a reliability coefficient within the limits of 0 and +1, since there is little chance for $r_{(S-1)(S-1)}$, the reliability of a composite, to become smaller than $r^2_{1(S-1)}$. Even, if $\bar{r}_{1P}$ should happen to be negative, r_{1t} would still be positive. This is not at all unreasonable since a reliability coefficient is a correlation between any two comparable series, and it is a mathematical impossibility for a large number of comparable series to have a consistent negative correlation with one another.

It should also be noted that when $a = 2$ the formula will apparently give a reliability coefficient of 0/0. Evaluating this by replacing the numerator and denominator with their derivatives, we come back to $\bar{r}_{1P}$. In other words, in the absence of a third judge corroborating the two, we can only take their intercorrelation as an approximation to the reliability of either. The effect of pulling here cannot be corrected since it is impossible to ascertain the direction of the pull.

Compared with the more elaborate method previously described, this new formula has distinct advantages on the side of practical procedure. All the intercorrelations need not be calculated. $\bar{r}_{Pq}$ and $\bar{r}_{1p}$ can be derived by means of Kelley's formulas (171) and (174) after calculating the correlation of each of the several series with their sum. The labor thus saved is considerable.

For purposes of illustration, the reliability coefficients calculated by this formula for the 26 judges on eight different traits are given in the accompanying table, together with the average correlation of each judge with the others of the same group. They are designated by r_{1t} and $\bar{r}_{1P}$, respectively.

The table retains one figure less than the actual calculation, so that slightly different reliability coefficients may correspond to apparently same average correlations. These corrected reliability coefficients are more discriminative than the average correlations, and are larger or smaller than the latter according as they are above or below the average of the whole group and according as they have been pulled down or pulled up in the intercorrelations. It is hardly necessary to point out that the average of the reliability coefficients will be higher than the average intercorrelation, since each intercorrelation is a geometric mean of two reliability coefficients and a geometric mean is always smaller than the arithmetic.

TABLE SHOWING THE AVERAGE CORRELATION AND THE CORRECTED RELIABILITY COEFFICIENT

Trait	Intellectual quickness		Intellectual profoundness		Memory		Impulsiveness	
Judge	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}
A	.71	.73	.61	.56	.51	.68	.30	.26
B	.76	.86	.69	.76	.51	.68	.42	.59
C	.66	.61	.63	.60	.43	.44	.36	.40
D	.67	.65	.68	.72	.52	.69	.22	.13
E	.71	.73	.71	.80	.47	.55	.40	.51
F	.64	.56	.58	.49	.10	.02	.34	.34
G	.70	.72	.65	.66	.43	.44	.31	.27

Trait	Adaptability		Persistence		Leadership		Scholarship	
Judge	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}
H	.29	.51	.24	.20	.58	.51	.68	.65
I	.18	.15	.21	.16	.68	.74	.60	.49
J	.15	.11	.23	.19	.68	.75	.70	.70
K	.08	.02	.16	.08	.56	.48	.68	.66
L	.18	.16	.25	.22	.50	.36	.72	.75
M	.28	.49	.36	.58	.70	.81	.76	.85
N	.22	.26	.41	.81	.75	.96	.78	.90

Trait	Intellectual quickness		Memory		Adaptability		Leadership	
Judge	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}
O	.63	.61	.57	.54	.54	.56	.53	.36
P	.57	.48	.69	.68	.51	.46	.67	.65
Q	.62	.59	.36	.19	.51	.47	.68	.67
R	.64	.63	.61	.64	.60	.71	.70	.73
S	.74	.91	.70	.89	.57	.62	.77	.94
T	.67	.70	.64	.71	.49	.43	.75	.86

Trait	Intellectual profoundness		Impulsiveness		Persistence		Scholarship	
Judge	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}	$\bar{r}_{1p}$	r_{1i}
U	.59	.56	.39	.63	.56	.74	.57	.43
V	.66	.75	.22	.15	.48	.49	.76	.87
W	.51	.40	.23	.17	.52	.60	.67	.64
X	.69	.86	.24	.17	.52	.61	.72	.77
Y	.53	.43	.32	.37	.42	.35	.69	.69
Z	.66	.75	.36	.48	.32	.19	.76	.86

AN OBJECTIVE TEST IN EDUCATIONAL PSYCHOLOGY[1]

W. S. MILLER

University of Minnesota

The problem of assigning marks in Educational Psychology to classes of 100 to 200 students each quarter led the author to formulate an objective test of 210 items which has been given without change for 10 consecutive quarters.

The test consists of "true-false" "multiple-choice," and "completion" items, totaling 210. It is possible for the slowest students to complete the test in two hours. The test contains some items easily enough passed by all members of the class and other items increasing in difficulty to the last one that is passed by only 10 per cent of the class.

It would be practically impossible to prepare each quarter such an extensive test, the items of which would differ from those of the preceding quarter. To avoid this difficulty it was planned to guard the test carefully and repeat it each quarter. Prior to the examination the mimeographed tests are numbered; each student gets the test paper corresponding to his seat number. All papers are carefully checked at the close of the examination. The tests are never used as teaching devices. Students are not permitted to have any other paper except the tests, thus preventing the copying of items of the test for further reference. Seldom are such precautions taken with the essay type of examination, and as a result, complete files of such examinations in many subjects for several years are available for students who may desire to use them.

If knowledge of the items in the objective test became common through coaching or through conversation about them, one would expect that through 10 repetitions of the test there would be an increase in the class medians. The following percentile distributions of scores show how little repetition has affected the results of the test in succeeding quarters.

This table should be read as follows: In the Fall Quarter 1922, thirty per cent of the class made less than 99 points on the test.

[1] The author is indebted to Herbert Sorenson, North Western University and William P. Reed, Aurora, Minn., for much of the statistical work in this investigation.

TABLE I.—PERCENTILE DISTRIBUTION OF SCORES ON OBJECTIVE TEST IN EDUCATIONAL PSYCHOLOGY, COLLEGE OF EDUCATION JUNIORS AND SENIORS

Quarter	Number	Lowest score	Percentiles									Highest score
			10	20	30	40	50	60	70	80	90	
Fall 1922	140	51	80	91.7	99	104.6	112.2	118.5	123.1	130.5	138.3	160
Winter 1923	61	62	75.5	96	102.6	110.3	115.4	119.7	129.5	139.5	146.1	159
Spring 1923	83	51	80.3	88.6	94.9	99.6	106.2	112.1	117.6	124.5	133.9	167
Summer 1923	181	41	85.9	95.1	102	107.5	113.1	121	127.4	135.3	144.9	167
Fall 1923	170	22	79.3	95.5	102.1	107.3	113.1	120	125.6	131.1	140.7	189
Winter 1924	75	55	80.4	90	98.5	103.8	111.9	117.5	125.5	128	133.8	164
Spring 1924	121	62	80.8	90.2	99.6	105.4	112.9	120.2	123.3	127.4	133.3	173
Summer 1924	261	52	85.04	95.76	103.2	110.1	116.8	123.8	129.7	136.7	149.2	169
Total	1092	22	81.3	93.3	100.8	106.8	113.23	120	125.75	132	141.97	189

The test has been given as a final examination for 10 quarters. The results of the two halves of the summer session are combined into a single group since the class in the second half is small.

In the Winter Quarter 1922–23 the test was given on the opening day and also on the last day.

TABLE II.—PERCENTILE DISTRIBUTIONS OF SCORES FOR THE FIRST AND LAST DAY OF THE COURSE

	Number	Lowest score	Percentiles									Highest score
			10	20	30	40	50	60	70	80	90	
First day	56	−5	23	29	35.2	40.4	45	49	52	58	79.2	91
Last day	56	62	80	100	104	110.8	117	120	130.2	139.4	146.4	159
Ten quarters	1054	22	81.23	93.28	100.7	106.81	113.28	120.02	125.81	132.18	142.09	189

The data presented in Table II show clearly that there is a gain in information which the test is designed to measure. A comparison with the results of the test in other quarters shows that practically all, if not all, of this gain is due to instruction during the quarter and not merely to the repetition of the test. This repetition of the test may account for the relatively high scores in the final for that quarter. One cannot be certain, however, since in the summer of 1924 still higher scores were made in the final, although the test was not given on

the opening day. The results in the two summer sessions are uniformly better than the average for the 8 quarters even though on the basis of intelligence tests the summer session groups show no selection upward.

About 18 per cent of the class did better on the first day than the poorest student did on the last day. On the last day about 82 per cent of the group did better than the best student on the first day. Scores below 80 on the last day were not given credit in the course. Nine per cent of the group made scores above 80 on the first day.

All of the students taking Educational Psychology had had 6 quarter credits or their equivalent in General Psychology and some, coming from normal schools, had had a course in Educational Psychology.

The correlation between the scores on the first day and the last day was .61 ± .056.

Except for the Spring Quarter 1923, when Army Alpha was given, and for the Winter Quarter when only Form A was given the students were given Form A and Form B of the Miller Mental Ability Test. The percentile distributions of the average of the two forms for each quarter and for the total are given in Table III.

TABLE III.—PERCENTILE DISTRIBUTION OF SCORES ON THE AVERAGE OF FORM A AND FORM B, MILLER MENTAL ABILITY TEST, COLLEGE OF EDUCATION JUNIORS AND SENIORS

Quarter	Number	Lowest score	Percentiles									Highest score	Tests
			10	20	30	40	50	60	70	80	90		
Fall 1922.....	140	66	78.3	85.4	90.8	94.7	97.6	100.3	102.8	105.2	108.8	114	Average of Miller A and B
Winter 1923[1].....	61	41	67.8	80.3	86.8	90.2	93.25	96.1	98.6	103.5	108.1	112	Miller A only
Summer 1923....	181	54	72.9	85.2	88.7	92.1	95.4	98.7	101.4	103.8	107.3	114	Average of Miller A and B
Fall 1923.....	170	51	77	83.7	88.4	92	95	98.7	101.3	103.4	106.2	116	Average of Miller A and B
Winter 1924.....	74	46	75.4	84.7	90.6	92.4	94.7	97.7	100.5	102.6	104.6	109	Average of Miller A and B
Spring 1924.....	118	58	81.1	84.6	88.1	91.2	93.9	96.5	99	101.9	105.1	110	Average A and B
Summer 1924....	254	43	70.7	81.3	86.4	90.9	94.3	98.5	101.8	104.6	108.3	117	Average A and B
Totals.....	998	41	74.8	83.6	88.2	91.9	95	98.3	101.2	103.8	107.4	117	S.D. = 12.57 M. = 93.2

[1] Only Form A was given to this group.

Had both Form A and Form B been given in the winter of 1922–23 the lower half of the group would have been a few points higher and the upper half a few points lower at each decile. The median would have been about 1 point higher.

The results show very little fluctuation in the median performance on the test for the 7 groups and no marked change in variability.

RELATION BETWEEN THE OBJECTIVE TEST AND ACADEMIC STANDING

We may now raise the question of the relation of the standing in the objective test to the academic achievement of the student as measured by the average honor points earned in the courses taken at the University. In calculating honor points, an A gives 3 honor points; B, 2 honor points; C, 1 honor point.

TABLE IV.—MEASURES OF CENTRAL TENDENCY AND VARIABILITY OF THE OBJECTIVE TEST, THE AVERAGE HONOR POINTS AND INTELLIGENCE TEST

	Juniors and seniors					Graduates				
	N	Md	M	SD	CV	N	Md	M	SD	CV
Objective test	1092	113.2	112.5	23.4	20.7	178	115.3	115	21.9	19
Average honor points	....	.....	.....	.....	....	178	1.66	1.686	.38	22.5
Average Miller Test A and B	998	95	93.2	12.57	13.5	139	97.3	94.5	10.9	11.5

The group designated "graduates" is made up of 178 students who completed the requirements for graduation (180 credits with 180 honor points) in June, 1924. One hundred thirty-nine of these graduating students had had Form A and Form B of the Miller Mental Ability Test. Among the remaining juniors and seniors are those who completed the graduating requirements in June, 1923. A few of them will, of course, be eliminated during the junior and senior years, and some will complete the graduating requirements in June, 1925. It should be emphasized that the graduates are a selected group, but not in the ratio of 178 to 1092.

It is quite evident from Table V that, barring the one exception, few students below C (92) in the test are above the mean (1.686) in honor points.

There are, however, 15 cases (upper left) who do strong B and A work in the test and who are *below* the mean of the group in honor points. There are two factors that are probably responsible in part for this discrepancy.

TABLE V.—RELATION BETWEEN OBJECTIVE TEST SCORES AND AVERAGE HONOR POINTS FOR 178 GRADUATES, COLLEGE OF EDUCATION, 1924

Objective test scores	.9	1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	f
180										1		1
170												0
160									1			1
150				1	3	1				1		6
140		1	2	1	1	5	2	2				14
130			3	7	6	3	2		1			22
120		2	3	7	5	4	3	2	2			28
110	1	5	8	13	4	2	3					36
100	2	6	5	12	4							29
90		5	4	4	2		1					16
80	3	6	3	5								17
70	1	1							r = .55 PE .035			3
60		2	1									3
50	1	1										2
f	9	29	29	50	25	15	11	4	4	2		178

Average honor points

1. The majority of these cases are majoring in mathematics and science, departments in which the central tendency of the marks is lower than the average for all departments, even though both the intelligence test and the objective test show that in ability and in achievement in educational psychology they are as a group better than the average of the 178 graduates.

2. Educational Psychology is a senior college subject and is therefore taken at a time when some students, especially those near the minimum "C" average, are making an effort to make up for low marks made in the junior college. The average honor points for the senior college is higher than the average honor points of the same students in the junior college.

In the group whose average honor points are relatively much higher than their standing in the objective test the majority are majoring in languages, physical education, and art education—departments whose average honor points are higher than the departments of science and mathematics referred to above.

There are, doubtless, many other factors that operate to produce the discrepancy between the test and average honor points. These factors are here enumerated although no evidence is available to prove that certain of them are operating in particular individuals.

1. Educational psychology is a required professional course for teachers which certainly would not call forth effort equal in all cases to that in the students' major interest.

2. Personality traits such as persistence, industry, cooperation, courtesy, general attitude, etc., are registered in the objective test only to the extent that such traits affect actual achievement. Few will deny that where the ratings are less objective such traits are more likely to affect the students' marks by making them higher than they would be if based on achievement only.

3. There is a possibility that some students do not do themselves justice on such an objective test; and, on the other hand, there are students who do not express themselves well in the essay type of examination, and are, therefore, favored by the objective form of examination. It is difficult for the instructor to divorce spelling, penmanship, and composition from knowledge of a subject. Indeed, instructors can be found who insist that these factors should all enter into the student's mark.

4. It would be expected that among 178 graduates there would be some students who, because of certain physical and mental stresses during this two hour examination, would fail to show achievement on a par with that represented by the average of four years of college work.

5. It would be unusual, indeed, if during the quarter they took Educational Psychology some of these students' programs of study were not disturbed by extra-curricular activities, such as dramatics, athletics, debate, etc., and by abnormal home conditions that would cause the performance in the test to be relatively lower than the standard indicated by their average honor points.

When one takes into consideration these various factors, and when one considers that the criterion with which the objective test is correlated is a composite of marks received in the various departments in which it is known that uniform standards of marking are exceptional, the correlation of $+.55 \pm .035$ seems quite significant and even more significant when one considers the degree of selection of the group.

It is apparent from Table VI that the graduates made better scores than the 1092 juniors and seniors.

TABLE VI.—PERCENTILE DISTRIBUTIONS OF THE SCORES ON THE OBJECTIVE TEST OF THE 178 GRADUATES AND THE 1092 JUNIORS AND SENIORS

	Number	Lowest score	Percentiles									Highest score
			10	20	30	40	50	60	70	80	90	
Juniors and seniors	1092	22	81.3	93.3	100.8	106.8	113.2	120	125.75	132	141.9	189
Graduates	189	58	85.8	96.6	104.3	111.25	115.3	120.3	126.6	136.2	143	189

RELATION BETWEEN INTELLIGENCE TEST SCORES AND THE OBJECTIVE TEST

In the first half of the summer session 1924, the 29 students making the lowest average scores on Form A and Form B of the Miller Mental Ability Test were placed in one section which we shall designate Section 3. The next 38 students in order were placed in Section 2 and the remaining 156 of a class of 223 were placed in Section 1. These sections were taught by different instructors. With 5 exceptions all of the 156 students in Section 1 made average scores on the Form A and B of the Miller Mental Test that would place them in the best two-thirds of the total group of 223 students.

TABLE VII.—PERCENTILE DISTRIBUTIONS OF SCORES MADE ON OBJECTIVE TEST BY SECTIONS 1, 2 AND 3

	Number	Lowest score	Percentiles									Highest score	M	SD	CV
			10	20	30	40	50	60	70	80	90				
Section 1	156	52	94	103	111	119	126	132	136	143	154	169	123.2	23.5	19
Section 2	38	64	85	92	98	106	112	115	118	124	132	146	110.1	17.5	15.9
Section 3	29	52	75	77	78	84	91	96	98	101	105	130	90.6	16.8	17.4
Total	223	52	85	95	103	110	117	124	131	137	150	169	116.8	24.5	21

More than 90 per cent of Section 1 exceeded the median of Section 3, and about 80 per cent of Section 2 exceeded the median of Section 3. More than 40 per cent of Section 1 exceeded the best score made by Section 3.

These large group differences in achievement exist in spite of the fact that in Section 1 the instruction was almost wholly by lecture with no oral or written quizzes except at mid-term, while Sections 2 and 3, being much smaller, were conducted on the recitation plan, with numerous oral and written quizzes.

In Section 1, five chapters of the text, Starch's "Educational Psychology," were assigned but not mentioned in the lecture, while these same chapters were all discussed in Sections 2 and 3.

In the more selected group of graduates, the relation between intelligence test score and objective test is not so close as it is in this large unselected summer class.

TABLE VIII.—RELATION BETWEEN MILLER MENTAL ABILITY TEST AVERAGE FORM A AND B, AND OBJECTIVE TEST, 139 GRADUATES

Average Miller A and B in tenths of SD Negative 5 SD = 0	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	f
65											2	3	2				1	8
60								1	1	3	3	5	2	4	2			21
55								2	4	3	8	7	3	1	1			29
50							1	2	5	7	8	5	2					30
45						2	1	2	3	2	6		1	4				21
40								1		2	2	3						8
35						1	1		2	3	3		1					11
30								1	1	2		1						5
25							1						1					2
20								2		1								3
15																		0
10																		0
5										1								1
0																		0
f						3	4	11	16	24	32	24	12	9	3	0	1	139

$r = .37 \pm 05$

Objective test in tenths of SD
5SD negative = 0

In Table VIII, the scores are translated into tenths of SD of the entire groups reported in Tables I and III with zero at negative 5SD.

From this scattergraph, Table VI, it is very apparent that in such a selected group achievement in the objective test is conditioned by factors other than the ability to perform or the Miller Mental Ability Test. The reasons for the relatively low correlation, $+.37 \pm .05$, need not be discussed here, since there is no experimental evidence in this study to prove what part industry, study habits, extra curricular activities, unreliability of tests, etc., play in reducing the correlation.

Whatever the factors may be they seem to operate in much the same degree in reducing the correlation between intelligence test scores and average honor points of the same students. Here the correlation is $+.385 \pm .049$.

TABLE IX.—RELATION BETWEEN MILLER MENTAL ABILITY TEST FORM A AND B WITH AVERAGE HONOR POINTS, 139 GRADUATES

Average Miller A and B in Tenths of SD Negative 5SD = 0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	f
65									3	3		1				1	8
60								1	4	5	2	3	4		2		21
55						1	1	3	7	4	2	5	3	1		2	29
50						2	3	10	8	3	1	3					30
45						2	2	3	4	5	1	4					21
40						1				4		2			1		8
35						2		3	3	2		1					11
30							1		3	1							5
25						1		1									2
20						1	1		1								3
15																	0
10												$r = .385 + .049$					0
5						1											1
0																	0
f						11	8	21	33	27	6	19	7	1	3	3	139

Average Honor Points in Tenths of SD
5SD negative = 0

SUMMARY

1. It is possible to construct a comprehensive objective test in Educational Psychology that will measure content of the course as evidenced by the increase of scores at the end of the course over those at the beginning of the course.

2. There is some overlapping of scores on the last day with those on the first day.

3. The correlation between scores on the first day and scores on the last day is $+.61 \pm .056$.

4. It is possible by taking the precautions mentioned to use the same test for 10 quarters without any noticeable increase in central tendency or variability of scores.

5. The results of such a test correlate $+.55 \pm .035$ with average honor points earned by 139 graduates majoring in practically all departments represented in the secondary school curriculum.

6. The central tendency of scores on the Miller Mental Ability Test used in successive quarters shows no regular increase.

7. The sectioning of students on the basis of the Miller Mental Ability Test results in marked differences of the sections in their performance on the objective test. These differences exist in spite of the fact that the less capable sections are approximately only one-fifth as large as the best section, and that recitations and oral and written quizzes are present in the small sections and absent in the large section.

8. The correlation between Miller Mental Ability Test scores and the objective test scores of 139 graduates is $+.37 \pm .042$.

9. The correlation between the Miller Mental Ability Test scores and the average honor points of 139 graduates is $+.385 \pm .049$.

A CLASS EXPERIMENT ON INDIVIDUAL DIFFERENCES IN MEMORY

HARVEY A. PETERSON

Illinois State Normal University

Norsworthy, Pyle, Lyon, and others have studied the relation of the rate of learning to retentiveness. They agree that those who learn quickly retain a larger number of units than those who learn slowly, and in some cases, as large a proportion of what has been learned as those who learn slowly. It seemed to the writer that it was possible to organize an interesting and profitable class experiment on this topic based particularly on the researches of Lyon but using longer selections.[1]

To a normal school class of 56 students a geographical selection of 250 words was given in mimeograph form. It described the character of the Philippine Islands and people. The first five lines will suffice as a sample:

The Filipinos as a race are not vicious./Their common crimes are rather slight offenses, such as the theft of articles of small value./Gambling is perhaps the chief vice,/and a Filipino will risk his last penny on a cock fight/or a horse race./ Americans are apt to think the Filipinos are lazy./This is partly true./For centuries they have learn-[2]

The numbers and the oblique lines were not on the learning sheets, and are inserted here merely to aid in explaining the methods of grading employed later. The students were given two and one-half minutes in which to study the selection, and then reproduced as much of it as they could in their own language immediately and again a week later. The first is taken as a measure of what was learned; the second, as a measure of what was retained. The learning was with knowledge that the reproductions would be called for.

A division of the selection into 36 "facts" or ideas, as shown by the numbers and oblique lines in the sample above, was then used to find the ratio of the number of facts which each person retained to the number which he had learned. This is called the *percentage of*

[1] Lyon, D. A.: "The Relation of Quickness of Learning to Retentiveness."

[2] Wright, Hamilton M.: "Handbook of the Philippines." Chap. VI.

retention. The group was then arranged in the order of the number of units learned, and divided into quartiles. The results for the different quartiles follow. The figures in the second and third columns give the average number of facts learned and retained respectively. The fourth column gives the average ratios of retention to learning by quartiles.

RATIOS OF RETENTION TO LEARNING IN THE FIRST SELECTION

Rank of quarters	Average learning scores	Average retention scores	Average percentage retained	Standard deviation
1	25.5	22.2	88	3.6
2	19.1	15.3	80	6.0
3	14.9	12.7	85	5.3
4	9.1	7.9	88	5.5
Averages.....	17.1	14.5	85	5.1

In general the four quartiles retained about the same percentage of what they had learned, *viz.*, from 80 to 88 per cent. Inasmuch as the quartile which learned the most retained about two and one-half times as much as the quartile which learned the least, the advantage is clearly and decidedly with the more rapid learners. The differences in the percentages retained in the different quartiles are not significant, for repeating the experiment five times, each time with a new group, showed very small average deviations for each quartile from the average of all 24 ratios. The average of all the ratios is 82; the average deviations of the four quartiles from this average are plus 2.2, for the first quartile, minus 0.3 for the second, 0 for the third, and minus 0.8 for the fourth.

With the same group of subjects a much more difficult selection, "The Policies of Labor Unions," was next tried. It dealt with the theory of labor unionism, and contained about 900 words. The first 10 lines were:

Collective Bargaining. (1) Its purpose. If the whole body of workers of a given kind can be brought into the union, so that the union can meet the employers as the representative of the whole, the position of the worker will be greatly strengthened. The fear that if he refuses to accept certain terms, another man will be employed in his place is removed. His ignorance of the market conditions

will be partly remedied both through the combination of the knowledge of all the workers of the union, and in some cases by the broader outlook which the union officials, partly or wholly exempted from daily application to manual work, may be able to obtain. The whole matter of bargaining can be put into[1]

Two changes in method were made. First, on account of the greater length of the selection the subjects answered questions about it instead of reproducing it. Secondly, each subject determined his own length of learning time, knowing that it was an element in determining his score. The learning scores and retention scores of different subjects were made comparable by calculating the amount learned and the amount retained per minute of time invested in learning by each subject. A summary of the results follows. The figures in the second column give the average learning scores of the quartiles. If all the questions had been answered correctly, the score would have been 100 per cent. The figures in the third column give the average retention scores after one week; those in the fourth column the average learning times in minutes. The figures in the fifth and sixth columns give the average learning and retention scores per minute of time spent in learning (MLT), and those in the seventh column, the average percentages of retention per minute of learning time.

RATIOS OF RETENTION TO LEARNING IN THE SECOND SELECTION

Quarters	Average learning scores, per cent	Average retention scores, per cent	Average learning times	Average learning scores per MLT, per cent	Average retention scores per MLT, per cent	Average percentage retained, per cent	Standard deviations
1	74	68	9.5	7.8	7.1	91	11.5
2	57	47	10.5	5.4	4.5	83	11.7
3	40	37	12.2	3.3	3.1	94	25.3
4	21	20	12.9	1.6	1.5	94	28.4
Averages.....	48	43	11.3	4.5	4.0	90	19.2

The quartile which learned the most took the least time for study. So far as the quartile averages go, the higher the learning ability, the less time needed for study, *in spite of the superior gains from study.* With the increase in the difficulty of the material the range between the best and the poorest learners increases. Whereas in the first selection the fastest quartile learned two and one-half times as much

[1] Bullock, C. K.: "Readings in Economics," Chap. XIX.

as the slowest, in the second selection the fastest quartile learned nearly four times as much as the slowest. The selection was intentionally somewhat difficult for the class. An idea of how difficult it was may be gained from the fact that the average mark of the whole class in learning was 48 per cent, a perfect answering of all the questions being 100 per cent. We are most interested in the percentages of retention, however. The first, third, and fourth quartiles retain about the same percentages—91, 94, and 94; the second retains 83 per cent. The standard deviations are large in the two slowest quartiles because when a person learns very little, an ability to answer one question more or less than another person can answer greatly changes the ratio of retention to learning, if their original scores were the same.

The correlations between learning and retention based on the absolute numbers of units, *i.e.*, numbers of "facts" in the first selection, and percentile marks in the second selection, are very high. For the first selection the correlation is .87 $\pm$.02 and, for the second selection, .94 $\pm$ .01. Summarizing our results we may say that differences in the rate of learning between the best and the poorest in a group increases with increase in the difficulty of the subject matter. In this investigation at least it is true that when persons are allowed to determine the length of their learning times, those who learn the most take the least time. Concerning the relation of retention to learning, there is a decided tendency for fast and slow learners to retain about the same proportion of what they learned, and this is true whether they reproduce what they learned in their own language, or answer questions on it, and whether the subject-matter is easy or difficult.

The experiment has proved quite satisfactory as a laboratory exercise for the introductory course in Educational Psychology. Usually only one selection is used with any given class, and in that form it can be done in about two and one-half 60-minute hours. Each paper is graded by two persons. Numbers are used instead of the names of members of the class. Each student has as much time as he needs for writing his reproductions, but he keeps a record of the amount used for reference. The experiment fits in well with the texts most in use at present. If the method of answering questions on the selection is used, a printed answer key should be furnished the class, and a longer interval between learning and reproduction, say two weeks instead of one week, would probably give more satisfactory results.

EIDETIC IMAGERY AND THE BINET TEST

LEONARD CARMICHAEL

Princeton University

Through the courtesy of Professor E. R. Jaensch it was my privilege recently to observe, in the psychological laboratory of the University of Marburg, the performance of certain children who are gifted with what Dr. Jaensch has called "eidetic imagery."

Dr. Gordon Allport, who has recently presented an excellent summary of this new movement in genetic psychology, quotes the following characterization of the phenomenon:

"Among optical memory-images we find in addition to the customary 'visual image' an *Eidetic Image*. In the one case a former visual perception is merely 'imagined,' in the other case the object is actually 'seen.' A person may be able to remember clearly a visual experience and to describe it in detail without necessarily possessing eidetic images. The true eidetic image, in distinction to the visual memory-image, revives the earlier optical impression when the eyes are closed, in a dark room, and sometimes when the eyes are open, with hallucinatory clearness."[1]

The method used by the Marburg workers and by Dr. Allport in his investigations of eidetic imagery is simple. The child under observation is given a picture, often a complex silhouette, and after a short period of observation the picture is removed and the subject "projects" the eidetic image upon a gray mat held at reading distance. The accuracy of the report of a true *Eidetiker* is certainly remarkable.

Undoubtedly this phenomenon is worthy of more study, particularly in America where due to the influence of "behaviorism" the whole question of imagery is in process of critical re-evaluation.[2]

---

[1] Quoted by Allport, G. W.: Eidetic Imagery. *The British Journal of Psychology*, Vol. XV, Part 2, October, 1924, p. 100, from Urbantschitsch, *Über subjektive optische Anschauungsbilder*. For the suggested criteria for separating the eidetic image from the memory image and the after image cf. Allport, loc. cit. A summary of the movement is to be found in the works quoted by Allport. The best single source which is later than his references is: Jaensch, E. R.: *Über den Aufbau der Wahrenhmungswelt und ihre Structur in Jugendalter*, Leipzig, 1923.

[2] Cf., for example, Gates, A. I.: "The Psychology of Reading and Spelling," 1922, p. 99: "No one knows precisely what imagery is, on what mechanism it depends, or how to examine it accurately in children."

It is the purpose of the present note to call attention to the fact that those who may be interested in the problems of eidetic imagery will find in the "Binet Test" a ready tool for the selection for further study of those children who possess this special ability.

When an unusually good performance is given in response to Test X, 3 of the "Stanford Revision of the Binet-Simon Test" (drawing designs from memory), the presence of eidetic imagery may be suspected. In a similar manner, the tester may, after the report upon the pictures used in Tests III, 2; VII, 2; or XII, 7, as the case may be, *without further exposure* ask the child if he can still "see" the picture, and if so to tell what he observes. While the results secured in this way are not of great value in themselves, because of the complicating factor of previous report they appear, nevertheless, to be significant enough to show whether or not the child then under examination will be a profitable subject to study for the possession of eidetic imagery.

In this manner a preliminary report can be secured upon many children and the subjects for further investigation isolated in the course of routine examination, with but a very slight additional expenditure of time.

A STUDY OF CHILDREN'S ABILITY TO NOTE SIMILARITIES AND DIFFERENCES

VERNON A. JONES[1]

Director of Research, Public Schools, Richmond, Va.

Though many of the tests in the Stanford Revision of the Binet Scale have been criticized or commented upon for one reason or another, few writers have had anything to say about the tests involving the giving of similarities and differences. This article deals with some results obtained from an experiment conducted with these two types of tests on 415 children of Grades I to VI inclusive. The problems with which it deals are:

1. Does the ability to see and express differences between two objects, as represented by words or pictures, occur at an earlier age than the ability to note similarities, or vice versa?

2. If the ability to do one of these tasks occurs earlier in the mental development of a child, does this mean that that task remains easier throughout the elementary school?

3. Which of these tests correlates most highly with mental age?

As far as we know the tests of S and D[2] are properly placed in the Stanford-Binet Scale, but it would be impossible to conclude from the fact that on the average children of seven years MA can express a difference in the pairs of words given in this scale, whereas on the average children are at the eight-year level of MA before they can give similarities in another series of pairs, that therefore the ability to note similarities is characteristic of a higher level of mental development. The difference between the difficulty of the two tests may be due to the fact that one set of words is less familiar to the children tested than the other. Secondly, suppose some one should ask us to tell in what way *sour* and *sweet* are similar, that would be easy enough but the test would not be so easy if the questioner should then ask for a difference between the same two words. The difference in the difficulty of the two tasks just mentioned is due neither to a difference in familiarity

[1] For assistance in this investigation, the writer is indebted to the following especially: Dr. Rudolf Pintner, of Teachers College; Miss Rebecca Coffin of Lincoln School, New York City; Mr. J. E. Dransfield and his corps of teachers of Public School No. 3, West New York, N. J.; the teachers of Fulton School, Richmond, Va.; the teachers of Big Stone Gap, Va.

[2] Throughout this article D will be used to stand for differences, or the ability to see and express differences, and S for similarities.

with the words making up the pair (they are the same in both cases) nor to our being "born short" on giving differences. In this study it is desirable that the pairs of words and pictures composing one test be equally as familiar as those of the other, and that the pairs not be "loaded" so as to favor either S or D.

Method of Procedure.—The method was adopted of using the same pairs of words and pictures for both S and D in order to eliminate one variable. Moreover these pairs—52 pairs of words and 14 pairs of pictures—were submitted to 100 teachers who were asked to pass

TABLE I.—JUDGMENTS FOR WORDS AND PICTURES

	S harder than D	D harder than S
Words		
1. Dime—nickel	43	57
2. Dog—cat	58	42
3. Knife—saw	48	52
4. Airship—kite	53	47
5. Fork—knife	49	51
6. Orange—lemon	48	52
7. Book—newspaper	42	58
8. Ocean—river	51	49
9. Oyster—fish	59	41
10. Snow—ice	45	55
11. Banjo—cornet	50	50
12. Glue—cement	40	60
13. Music—noise	59	41
14. Latitude—longitude	57	43
	702	698
Pictures		
1. Fork—knife	43	57
2. Sun—electric light	41	59
3. Violin—phonograph	47	53
4. Parrot—chicken	50	50
5. Search-light—sun	51	49
6. Razor—ax	55	45
	287	313

judgment on each pair as to whether it would be easier for a child to see and express an S or a D. No equal judgments were allowed. The pairs of pictures were pasted in booklets, one pair to each page, so that only one pair was visible at the time. Out of the 52 pairs of words 14 were obtained where the ratios of judgments were not more one-sided than 40 to 60 or 60 to 40 that S was harder than D. In like manner 6 of the 14 pairs of pictures were selected. Table I gives the pairs of words and the judgments concerning each. Though the pictures cannot be reproduced here, an idea is given as to what they were.

These pairs of words and pictures were given individually to 96 children of Grades I to III inclusive, in Public School No. 3 West New York, N. J., Public School No. 43 New York City, and to 51 children of same grades in Lincoln School, New York. They were given as a group test to 268 Children of Grades IV to VI inclusive, of Public School No. 3 West New York, N. J. All the tests were given and scored by the writer in accordance with a uniform method. In both the individual and the group testing the pairs of words were dictated, and the pairs of pictures were presented one at a time by means of small booklets. Practically the only difference between the group and the individual procedures was that in the former the answers were written, in the latter oral. The tests for the two capacities were given separately. If, for example, the test for D was given first the individual or class went through all the words and pictures, giving only D, not knowing that S would be called for later. To equate as much as possible for the effect of mental set it was arranged that the cases where S was given before D alternate with the cases where D was given before S. The method of scoring was to count each S as 1 and each D as 1.[1]

Results.—In all the results presented, the data from Public Schools Nos. 3 and 43 will be combined, because the schools are made up of pupils of much the same type as far as the writer could discover. However the facts obtained from Lincoln School, where the pupils

[1] This method can be criticized due to the fact that all the pairs are not perfectly matched according to the judgments, and secondly no attempt has been made to scale the items of the tests. A more refined method was used on part of the data, but the results came out so near to those obtained by the crude method that the use of the more laborious procedure would not have been justified. And again the original data of this study are so crude that minute statistical treatment of them would be useless.

TABLE II.—SHOWING THE AVERAGE NUMBER OF DIFFERENCES AND SIMILARITIES GIVEN PER PUPIL FOR WORDS AND PICTURES IN GRADES I TO VI

		Words									Pictures						
	Grades	Schools 3 and 43		Number cases	Lincoln		Number cases	Total		Number cases	Schools 3 and 43		Lincoln		Total		Number cases
		D	S		D	S		D	S		D	S	D	S	D	S	
Individual tests	1	4.10	2.86	50	6.78	6.67	9	4.51	3.44	59	2.39	1.84	2.89	3.44	2.47	2.08	59
	2	4.46	4.07	41	6.87	8.48	23	5.31	5.66	64	2.41	3.07	3.96	5.39	2.97	3.89	64
	3	4.50	7.25	4	7.70	10.65	20	7.17	10.09	24	3.75	4.25	4.25	5.50	4.17	5.29	24
Group Tests		School No. 3									School No. 3						
	4	6.74	7.94	78							3.26	3.96					
	5	6.96	7.92	108							3.62	3.89					
	6	8.50	10.12	82							4.51	4.92					

examined were a highly selected group, require separate treatment. No children above Grade III were examined at School No. 43 and at Lincoln.

The outstanding points in the table are: First, the steady increase in the average score in S and D as you go up the school grades, second, the consistency between the scores in words and pictures, third, the consistency of all the figures in indicating that an S, as measured on this group of words and pictures, is harder to see and express than a D in the lower grades, and easier in the higher grades. Up to the end of the first grade in the case of superior children, and up to about the end of the second for the less gifted, S seems to be harder than D, though of course this depends finally upon mental age and not on school grade. It was possible to get mental ages for 110 of the 147 children of Grades I to III inclusive. Of those examined at Lincoln School Binet Mental ages were obtained for 39. Of the 95 tested at Schools Nos. 3 and 43, mental ages as determined by Pintner-Cunningham Primary Mental Test were secured for 71. In order to show general trends, mental ages as determined by both methods have been lumped together. For this group, therefore, it is possible to study

TABLE III.—SHOWING FOR VARIOUS AGE LEVELS WHAT PERCENTAGE THE D AND S RESPONSE, RESPECTIVELY, ARE OF THE TOTAL NUMBER OF CORRECT RESPONSES, RESULTS BEING GIVEN SEPARATELY FOR PICTURES AND WORDS

Age levels	Percentages				Number cases
	Pictures		Words		
	D	S	D	S	
Under 6^6	72	28	89	11	9
6^6–7^0	54	46	54	46	10
7^0–7^6	53	47	59	41	18
7^6–8^0	44	56	48	52	15
8^0–8^6	46	54	46	54	17
8^6–9^6	40	60	43	57	25
Above 9^6	44	56	44	56	16
					110

more directly the abilities to note S and D in comparison with mental age levels.

In order to discover whether or not any corrections are necessary for Tables II and III due to the fact of mental set—in what S was called for first in some cases and D in others—a comparison was made of the scores obtained under the two conditions.

TABLE IV.—SHOWING A COMPARISON BETWEEN THE AVERAGE NUMBER OF CORRECT RESPONSES PER PUPIL WHEN S WAS GIVEN FIRST AND WHEN D WAS GIVEN FIRST, FOR DIFFERENT GRADES AND SCHOOLS

			D First				S First				
	Grades	Schools	Number cases	Similar	Different	S—D	Number cases	Similar	Different	S—D	Net advantage due to being given first
Pictures	1	Nos. 3 and 43	26	1.77	2.50	− .73	24	1.96	2.25	− .29	+ .44
		Lincoln	5	4.40	3.60	+ .80	4	3.00	1.25	+1.75	+ .95
	2	Nos. 3 and 43	22	3.73	2.36	+1.37	19	2.32	2.47	− .15	−1.52
		Lincoln	12	5.25	3.42	+1.83	11	5.54	4.18	+1.36	− .47
	3	Nos. 3 and 43	2	4.00	3.50	+ .50	2	4.50	4.00	+ .50	.00
		Lincoln	8	5.38	4.63	+ .75	12	5.58	3.92	+1.66	+ .91
	4, 5, 6,	Nos. 3 and 43	113	4.00	3.30	+ .70	155	4.40	4.10	+ .30	− .40
Words	1	Nos. 3 and 43	26	3.04	4.27	−1.23	24	2.67	3.92	−1.25	− .02
		Lincoln	5	7.60	7.20	+ .40	4	5.50	6.25	− .75	−1.15
	2	Nos. 3 and 43	22	4.86	4.90	− .04	19	3.16	3.95	− .79	− .75
		Lincoln	12	8.93	6.92	+2.01	11	8.82	7.45	+1.37	− .64
	3	Nos. 3 and 43	2	7.00	4.00	+3.00	2	7.50	5.00	+2.50	− .50
		Lincoln	8	10.38	8.75	+1.63	12	10.75	6.92	+3.83	+2.20
	4, 5, 6,	Nos. 3 and 43	113	8.00	6.80	+1.20	115	9.10	7.80	+1.30	+ .10
		Totals	188				227				

The figures in the column marked "Net Advantage" are obtained by subtracting the figures of column S—D of "D First" section from S—D of "S First" section. If the mere giving of S or D first gives advantage to the one given first over the one given last, due to mental set working against the latter, then the figures of this column will be +, if not, 0 or —.

This table indicates that there is no consistent tendency for a lower score to be obtained on the test given last. This tendency is present in the results for Grade I for pictures, and, possibly, it can be interpreted as due to the interfering influence of the first test. How-

ever, the tendency is not found among the results on words for the same grade. The high scores found in the "Net Advantage" column for Grade III of Lincoln School are quite out of agreement with the other figures of the table, and though the writer has no explanation to offer for these, it should be noted that all the results for the Grade III are based on fewer cases that those for any other grade. Except for the two cases mentioned, the general trend on these tests is for children, not to be at an advantage, but often to be at a disadvantage, on the test given first. This does not mean that mental set plays a negligible part—it plays a considerable rôle as will be shown by Table VI—but it means that the disadvantage due to interference of the previous test may be equalled by the advantage of having become more familiar with the test and the examiner.

It was felt that some additional evidence on the relative difficulty of noting S and D for these words and pictures at different age levels could be obtained by recording the number of cases in which children give an S when called upon for a D, and vice versa. Table 5 gives the data concerning these reversals for the 147 children examined individually.

TABLE V.—SHOWING THE AVERAGE NUMBER OF REVERSALS PER CHILD FOR D AND S AT VARIOUS GRADE LEVELS ON PICTURES AND WORDS

Grade	School	PD	PS	WD	WS	Number cases
First	3 and 43	.44[1]	1.34	.74	2.56	50
	Lincoln	0	.33	.22	1.22	9
	Combined	.37	1.19	.67	2.36	59
Second	3 and 43	.10	.61	.15	1.39	41
	Lincoln	.09	.13	.0	.30	23
	Combined	.09	.44	.09	1.00	64
Third	Combined	0	0	0	0	24

[1] The figure .44 means that in the first grade of Schools Nos. 3 and 43 when 6 pairs of pictures are given and a D asked for, each child, on the average, gave an S instead of a D in .44 pairs.

Care must be exercised in the interpretation of the facts of Table V. Whether the wrong response is looked for as a result of not understanding the task, or whether the wrong response—suggested to the

subject during the delay in which he is looking for the right response—is accepted as satisfactory due to a weakness in his concept, it is difficult to say. But whatever the explanation be, it is significant that the reversals in the S columns are, in every instance, higher than those in the D columns, and that they decrease rapidly as we move up the grades and to higher mental ages.

TABLE VI.—AVERAGE NUMBER OF REVERSALS PER PUPIL WHERE S IS GIVEN FIRST, AND WHERE D IS GIVEN FIRST

	S First	D First
PD	.33	.08
PS	.63	.68
WD	.66	.04
WS	1.47	1.26
No. cases	67	80

A correlation was worked between the scores received on S and mental ages, as determined by the Stanford Revision of the Binet Scale, for 39 children of Lincoln School. The correlation was also worked between D and Binet Mental Ages. The r (Pearson) between S and mental ages was $+.67 +.06$, while the r between D and mental ages for the same group equalled $.44 +.09$. The range of mental ages was from 86 to 134 months. It is needless to say that we must be careful in concluding from these figures what the relationship will be at other ages, but it is especially doubtful if these figures are typical of what would be found if lower mental ages were used, such, for example, as 6 to 7.

Except for the lower age levels, it is believed that the correlations obtained on this small group are fairly typical of those that would be found with any normal group of elementary school pupils. In correlating test item 8, similarities, of the Otis Advanced Examination, Form A, with the median mental age derived from four intelligence tests,[1] the r was found to be $.69 \pm .03$ for 101 cases of Grades V to VIII inclusive.

[1] National A, National B, Haggerty 2 and Otis Adv. Examination.

THE RELATION OF SEX DIFFERENCES TO THE KIND OF MATERIAL USED

W. H. PYLE

University of Missouri

In all or nearly all forms of learning, girls excel boys. In ideational learning, however, the result depends in some measure upon the kind of material used. With some kinds of material girls excel; with other kinds, boys excel; and with some material, there is little or no difference. In the accompanying table are shown the results of four different kinds of ideational tests given to the same high school pupils. The tests are designated as A, B, C, and D. The methods of administering the tests were as follows:

In Tests A and B each pupil had a copy of the test which he studied for five minutes. The tests were then collected and questions on the material studied were given to the pupils. They were allowed five minutes to answer the questions. In Test C the procedure was the same except that the results of the pupils' learning were determined by the Ebbinghaus Completion Method. Test D was read once to the pupils who were then required to answer questions upon the material read. The tests were given to the same high school pupils, 288 girls and 196 boys, nearly equally distributed for the different ages.

TABLE I.—AVERAGE SCORES MADE ON TESTS A, B, C, AND D BY HIGH SCHOOL BOYS AND GIRLS

Age	Sex	Test A average score	Test B average score	Test C average score	Test D average score
14	Girls	9.0	6.5	25.5	19.2
	Boys	10.0	6.5	22.0	20.9
15	Girls	9.6	6.6	25.3	19.8
	Boys	10.6	7.8	22.1	19.8
16	Girls	10.5	8.6	27.1	20.9
	Boys	11.3	8.0	22.3	21.1
17	Girls	11.2	9.1	26.2	19.9
	Boys	14.1	11.0	25.7	19.7
Average.....	Girls	10.08	7.70	26.00	19.95
	Boys	11.50	8.33	23.03	20.38

A study of the table shows that in Test A the boys are better for all four of the ages, on an average excelling the girls by 14 per cent. In Test B at age 14, the boys and girls make the same score. At age 16, the girls make the better score. At ages 15 and 17, the boys make the better score. If we average the four years, it gives the boys the better score by 8 per cent. In Test C the girls are better for each of the four ages. In Test D the boys and girls make the same score at age 15; at ages 14 and 16 the boys are slightly better; at age 17 the girls make the better score. If the scores in Test D for the four ages are averaged, the boys have a better score by 2 per cent.

Copies of the four tests, A, B, C, and D are given below. It will be seen that Tests A and B are of a scientific nature. In tests of this type it seems in general that high school boys do better than high school girls.

In Test C, as the table shows, the girls are better at every age, making an average of 13 per cent better than the boys. Test C is about furs and the animals from which they come. One would think on *a priori* grounds that parts of this test would make a stronger appeal to boys; namely paragraphs 2 and 3, while paragraphs 1 and 4 would appeal more strongly to girls. The facts are, however, that in the test as a whole the girls very much excel the boys.

Test D is largely a matter of mere rote memory and in this test the sexes make about the same score.

The theoretical implication of this study is evident. If one had given only Test A or B to high school pupils, he would say that boys are better learners. If one had given only Test C, he would say that girls are the better learners; while if one had given only Test D, he would say that there is little difference in the learning capacity of high school boys as compared to that of high school girls.

Of course it is also quite possible that the method of testing efficiency had something to do with the results. Test A, B, and C were visual tests, while test D was auditory and presented but once. The efficiency was determined in A, B and D by means of questions and answers, while in C, it was determined by the Ebbinghaus method. The one point we wish to make here is that the material and probably the method have much to do with the results of ideational learning tests. Copies of the material used are given here for the use of others who may wish to use the same material.

A

Practically everything you eat, taste, wear, smell, and see, has resulted in some way from the ingenuity of chemists. The story of chemistry is like an endless chain—it might begin anywhere, and need never end.

Just now you were looking out of the window. That glass is a product of chemistry. Glass is made of soda, lime and sand. A mixture of these substances is melted down to a bright red heat. A big molten ball of it is then gathered on the end of a blowpipe. Air is forced through the pipe and the ball becomes a bulb, the bulb becomes a long cylinder. At the right temperature, the cylinder is laid on a table and slit. The cylinder flattens out. That is window glass.

In the hall door of your home or in your office there is a pane of plate glass. This is made by casting the molten glass. It is first pressed as smooth as possible, then ground still smoother, then polished. It is a product of comparatively recent years, chiefly due to American ingenuity.

The desk at which you work was made with steel tools. The steel of which these tools were made is, of course, a chemical composition. And these wood-cutting tools had first to be made with other tools that could *cut* steel.

Once this process was very slow and expensive. The steel tool used by a lathe worker in cutting out other tools would get red-hot and lose its "temper," so that it could not cut. Then the lathe worker lost his temper, too! For he had to stop, resharpen his tools, and waste a great deal of time. Chemists added certain rare metals, such as tungsten or molybdenum to the iron that makes ordinary steel, and as a result of that and of other discoveries, we have "high-speed steel."

This steel is so expensive that tools are tipped with only a fraction of an inch of it in much the same way that our finest gold pens are tipped with a platinum iridium alloy to keep them from wearing out.

This steel tip is many times as efficient as ordinary steel. Without high-speed steel you could not have such cheap typewriters, farm implements, and automobiles, for this product has revolutionized the whole metal industry.

B

In the room where you are reading there is probably a tungsten electric light bulb. Doubtless you remember what a big improvement the first tungsten filament seemed to be, only a few years ago, when it came to displace the carbon filament in our electric bulbs. The carbon filament was inefficient and burned up too much power for the light it gave. So chemists had to seek a new substance that could be raised to a higher temperature, receive a higher current, and so give more light.

After trying various substances, they tried tungsten. This is a metal which we get as an oxide, like a powder, just as we get iron. The first difficulty was to make tungsten into a workable metal. This was solved by pressing the tungsten powder together with a "binder," just as a confectioner fastens popcorn balls together with gum arabic.

You remember how frail the filaments made of this tungsten metal were at first? If you had examined one of these filaments under a microscope, you would have seen that it looked something like a string of peas fastened together with the gluelike binder. A very slight jar was sufficient to break the filament.

So chemists set to work to make tungsten into a ductile metal that would bend and could be drawn out without breaking. Here they met a great difficulty, because the tungsten would form into crystal-like parts that were so brittle that the metal chipped instead of bending. Finally it was made ductile by the process of "swaging." Electricity is sent through a bar of tungsten until the bar becomes white-hot. Then the bar is passed through a rapid hammering process, with hammers beating the white-hot metal from all directions. This breaks down the crystals, and the bar of tungsten becomes a ductile metal.

After swaging, the tungsten is in the form of a metal bar about 14 inches long and three-eighths of an inch square. To get it into a filament for your electric light bulb, it has to be passed through a series of dies. After passing through each die, the bar is smaller around and longer, until finally it is drawn out into a wire. When it has passed through the last tiny die the original 14-inch bar has become a wire *thirty-nine and six-tenths miles* in length. In an ordinary electric light bulb, there are about 16 inches of this wire.

C

The rich fashionable woman that puts on her expensive furs on a cold winter day knows all about them from the point of view of their beauty, warmth and price, but knows little of their origin.

When Nature filled the cold northern lands with our four-footed brothers, she clothed them with coats that were warm, light, durable, and of protective color. These coats were thick in winter and thin in summer, and were made of fur, with its dense, woolly under coat and its long, soft, shiny outer coat—the inner for warmth, and the outer for wet and wear. So it turns out that the coat which Nature first made for a wild animal, later serves as a coat for a woman.

Some northern animals can store up food in holes or in the fat of their bodies, and need not be out in the cold all winter. Such animals as the Marten and Northern Fox that have to be out all winter have the best fur. Their fur is the densest, finest, lightest and therefore the softest and warmest of all furs.

The choicest colors of furs are soft olive brown with silver hairs found on the Russian Sable, and glossy black with silver hairs found on the true Silver Fox of the North.

[D]

The fabled island of Labdamon is in the Pacific Ocean about 38 degrees south of the equator. It was discovered in 1639 by a French sailor. It contains 50 thousand square miles and has a population of two million natives and 100,000 Spanish settlers, but no French. The people raise wheat but no corn, grapes but no oranges, turnips but no potatoes, cattle but no hogs. The island is noted for its small lakes and large rivers. The climate is relatively warm in winter and cool in summer. The amount of rainfall is great in the fall but little in the spring.

A NOTE ON STATISTICAL METHOD

C. L. HUFFAKER

University of Arizona

In certain types of investigations it is necessary to find a number, often a large number, of intercorrelations. The customary procedure is to make a frequency distribution for each test, and a scatter-gram for each correlation. The purpose of the frequency distributions is to check the accuracy with which the scatter-grams are plotted. By using the means and standard deviations of each distribution it is possible to find the intercorrelations without the plotting of scatter-grams, with a considerable saving of labor. The derivation of the formula is as follows:

Kelly gives the following difference formula for r using gross scores,

$$r = \frac{\Sigma_1^2 + \Sigma_2^2 - 2M_1M_2 - \Sigma^2 d}{2\sqrt{\Sigma_1^2 - M_1^2}\sqrt{\Sigma_2^2 - M_2^2}} \tag{1}$$

In this formula:

Σ_1^2 = Standard deviation squared of X_1 scores about O as an origin.
Σ_2^2 = Standard deviation squared of X_2 scores about O as an origin.
Σ_d^2 = Standard deviation squared of $(X_1 - X_2)$ about O as an origin.
M_1 = Mean of X_1 scores.
M_2 = Mean of X_2 scores.

$$\Sigma_1^2 = \frac{\Sigma X_1^2}{N} \quad \Sigma_2^2 = \frac{\Sigma X_2^2}{N} \quad \Sigma_d^2 = \frac{\Sigma(X_1 - X_2)^2}{N} \text{ (by definition).}$$

It has been shown[1] that the relationship of the standard deviation about the mean and about any other point is

$$s^2 = \sigma^2 + \delta^2$$

In the present case $s^2 = \Sigma^2$ and $\delta^2 = M^2$. Therefore,

$$\Sigma_1^2 = \sigma_1^2 + M_1^2 \text{ and } \Sigma_2^2 = \sigma_2^2 + M_2^2$$

Likewise, $\sqrt{\Sigma_1^2 - M_1^2} = \sigma_1$ and $\sqrt{\Sigma_2^2 - M_2^2} = \sigma_2$

Substituting in Formula (1) we have

$$r = \frac{\sigma_1^2 + M_1^2 + \sigma_2^2 + M_2^2 - 2M_1M_2 - \dfrac{\Sigma(X_1 - X_2)^2}{N}}{2\sigma_1\sigma_2} \tag{2}$$

[1] Kelley, Truman L.: "Statistical Method," p. 180.

Collecting terms

$$r = \frac{\sigma_1^2 + \sigma_2^2 + (M_1 - M_2)^2 - \dfrac{\Sigma(X_1 - X_2)^2}{N}}{2\sigma_1\sigma_2}$$

All of the constants entering into this formula except $\frac{\Sigma(X_1 - X_2)^2}{N}$ can be obtained from the frequency distributions. By using a table of squares and a calculating machine the term $\frac{\Sigma(X_1 - X_2)^2}{N}$ can be readily found.

A STUDY OF VARIABILITY IN ARITHMETIC

JOHN R. CLARK AND E. LEONA VINCENT

The Lincoln School of Teachers College

I. Variability in Single Column Addition

Purpose.—To study the variability of 87 Grade V and VI children in speed and accuracy in addition on 22 successive school days, four minutes being devoted to the work each day.

Materials.—In many studies of variability the materials used have been of so complex a nature as to make it uncertain whether differences in scores on successive trials were due to variability of pupils or to unreliability of test materials. It was believed that the factor of unreliability could be avoided in this study by the use of very simple materials. For this reason sets of single column additions of nine digits (no digit less than four repeated) were used.

Procedure.—The pupils of Grades V and VI of the Lincoln School, 87 in number, added single columns for four-minute periods on each of 22 successive school days. The periods were placed as nearly as possible at the same hour each day.

A study was made of the number of columns attempted and of the number right (*a*) for the group for each day and (*b*) for each child from day to day. The measures of variability used in the study of the individual children were:

1. Semi-interquartile range $\frac{Q_3 - Q_1}{2}$.
2. The median of variations from Q_2, *i.e.*, median of (measures $-Q_2$) here called V.
3. This median expressed in terms of the semi-interquartile range, $\frac{V}{Q}$.

These measures were computed for each child on the basis of his 22 daily performances, and were also studied from the point of view of their relation to mental age and intelligence quotient.[1]

Results.—Table 1 shows the medians of the number of columns attempted and of the number right, for each day.

---

[1] Measured by Stanford-Binet and adjusted to a date at the beginning of the experiment.

TABLE I.—SHOWING MEDIAN ATTEMPTS AND RIGHTS

Day	Median of attempts	Median of rights
First	17.0	14.5
Second	12.1	10.4
Third	12.8	10.6
Fourth	13.3	11.7
Fifth	13.4	12.3
Sixth	12.5	10.2
Seventh	15.3	13.3
Eighth	14.5	12.0
Ninth	14.8	13.0
Tenth	17.1	14.8
Eleventh	14.5	11
Twelfth	14.8	11.6
Thirteenth	19.5	15.9
Fourteenth	16.7	12.5
Fifteenth	15.0	12.5
Sixteenth	16.6	12.8
Seventeenth	17.2	14.0
Eighteenth	14.8	11.0
Nineteenth	17.0	13.6
Twentieth	16.7	14.0
Twenty-first	14.5	12.7
Twenty-second	18.4	16.2

The median number of examples attempted in the daily practice varied from 12.1 to 19.5 or through a range of seven examples. The semi-interquartile range of the medians was 2.02 examples. This means that the chances are even that the medians of the number of additions of this sort attempted by a group of pupils of this age will vary when they work under pressure of time by two or more examples.

Table 1 shows, too, that the median of the number of rights in the daily practice varied from 10.2 to 16.21 or through a range of six examples. The semi-interquartile range of the medians of the number right was 1.3, meaning that the chances are even that the medians of the number of correct additions of this sort for a group of pupils of this age will vary by 1.3 or more.

Table II shows the three measures of variability for the number of attempts and the number of rights. These figures, it will be recalled, were computed not on the basis of group medians, but of

individual performances. From the median of the individual semi-interquartile ranges we see that in a standardized test of four minutes of simple addition the chances are even that any child on any day may vary from his own median performance by 2.1 columns or more in speed and by 2.5 columns or more in number right.

TABLE II.—THREE MEASURES OF VARIABILITY FOR NUMBER OF COLUMNS ATTEMPTED AND NUMBER RIGHT

	Number columns attempted	Number columns right
Median of individual median Q^{10}	2.08	2.47
Median of individual median $V's$	2.11	2.18
Median of individual median V/Q	.96	1.00

The writers are at a loss to account for the fact that among individuals the variability in number of rights is greater than in number of attempts while the reverse is true for groups.

Relation between Variability and Intelligence.—Table III*A* shows a positive relationship between variability in speed and mental age. There is also a positive relationship between variability in speed and intelligence quotient as shown in Table III*B*.

TABLE III*A*.—SHOWING RELATION BETWEEN THE THREE MEASURES OF VARIABILITY IN NUMBER OF COLUMNS ATTEMPTED AND MENTAL AGE

Mental age	Number cases	Q	$\frac{V}{Q}$	V
16–0 up	5	2.8	1.30	3.40
15–0 to 15–11	3	2.9	.97	2.56
14–0 to 14–11	10	1.8	1.17	2.14
13–0 to 13–11	20	2.4	.99	2.40
12–0 to 12–11	22	2.7	.94	2.47
11–0 to 11–11	20	2.2	.93	2.11
10–0 to 10–11	6	1.7	.90	1.47
9–0 to 9–11	1	1.0	1.30	1.30

TABLE III*B*.—SHOWING RELATION BETWEEN VARIABILITY AND INTELLIGENCE QUOTIENTS

Intelligence quotient	Number cases	Q	$\frac{V}{Q}$	V
150–up	4	2.6	1.1	2.9
140–149	6	2.4	.9	2.0
130–139	8	2.2	1.1	2.4
120–129	23	2.8	1.0	2.8
110–119	23	2.3	.9	2.1
100–109	16	1.8	1.0	1.8
99 down	7	2.2	.9	2.0

TABLE IV*A*.—RELATION BETWEEN THE THREE MEASURES OF VARIABILITY IN NUMBER OF COLUMNS RIGHT AND MENTAL AGE

Mental age	Number cases	Q	$\frac{V}{Q}$	V
16–0 up	5	2.4	1.00	2.4
15–0 to 15–11	3	2.6	.90	2.4
14–0 to 14–11	10	2.1	.96	2.2
13–0 to 13–11	20	2.5	.97	2.4
12–0 to 12–11	22	3.2	.92	2.5
11–0 to 11–11	20	2.3	.96	2.0
10–0 to 10–11	6	2.1	.93	1.9
9.0 to 9–11	1	1.0	1.10	1.1

TABLE IV*B*.—SHOWING RELATION BETWEEN VARIABILITY AND INTELLIGENCE QUOTIENTS

Intelligence quotient	Number cases	Q	$\frac{V}{Q}$	V
150 up	4	2.6	1.0	2.6
140–149	6	2.6	1.0	2.6
130–139	8	2.3	1.0	2.3
120–129	23	2.6	.9	2.4
110–119	23	2.3	1.0	2.3
100–109	16	2.0	.9	1.8
99–down	7	2.8	.9	2.5

Table IV*A* and IV*B* show a somewhat smaller positive relationship between variability in number right and intelligence quotient, and a still smaller, almost negligible, positive relationship between variability in number right and mental age. This means that the people with high intelligence quotients and mental ages have a tendency to vary more from day to day in both speed and accuracy of addition than do those with lower mental ages and intelligence quotients. This we would expect, since the people with greater intelligence tend to do more work and hence have more room for variation.

II. Variability in the Four Forms of the Woody-McCall Mixed Fundamentals

Purpose.—To study the variability of 116 Grade V, VI, and VII children on the four forms of the Woody-McCall Mixed Fundamentals given on four successive days, one form each day.

Materials.—Forms I, II, III, and IV of the Woody-McCall Tests in Mixed Fundamentals. The data from this part of the study are complicated by the factor of unreliability of material, since it is impossible that four forms of a test covering all of the fundamental processes through so wide a range of difficulty could be absolutely reliable. It is a common practice in experimental work, however, to give one form of such a test, then to proceed with an experiment or a period of teaching, and finally to measure the result of the experiment or teaching by giving another form of the initial test and computing the differences in scores. The writers of this paper found in their work in addition such variability from day to day in the performance of children that it was thought desirable to determine the amount of variability that would be found in successive applications of a test of this sort without the interference of experimental or teaching factors.

Procedure.—The tests were given at the same hour by the same examiners on four successive days, the four forms being alternated in the five groups (two fifths, two sixths, and one seventh grade) so as to equalize the effects of practice. During each test the children were asked at the end of the first five minutes to draw a ring around the example they were working on at that time. It was thought possible in this way to get a measure of the variability in number of rights under an "active" time limit.

Results

Variability in Measures of Central Tendency.—Table V*A* shows the medians for the four forms of the test for each of the three grades. In

contrast to the medians of the single column additions in the first part of the paper we find these in close agreement from day to day. This illustrates the fact that measures of central tendency are generally characterized by stability, whereas measures of individual performances are extremely variable. The variability of the medians in Part I of this study is in all probability due to the functioning of an active time limit.

TABLE V*A*.—SHOWING THE MEDIAN SCORES FOR THE FOUR FORMS OF THE WOODY-MCCALL MIXED FUNDAMENTALS FOR EACH GRADE

	Grade V	Grade VI	Grade VII
Form I	21.19	25.80	27.00
Form II	20.50	25.31	26.60
Form III	20.92	26.00	27.75
Form IV	20.31	25.83	27.00

TABLE V*B*.—SHOWING THE MEDIAN NUMBER OF RIGHTS FOR THE FOUR FORMS FOR EACH GRADE AT THE END OF FIVE MINUTES OF PRACTICE

	Grade V	Grade VI	Grade VII
Form I	21.0	24.8	20.3
Form II	17.0	19.8	21.5
Form III	18.3	20.7	22.8
Form IV	20.5	20.3	Data missing

It should be remembered that the Woody-McCall Mixed Fundamentals is a power test rather than a test of speed. The time limit is not "active" since most pupils finish all they can do before the expiration of the time allowed. If, however, time is allowed to enter as an active factor, variability of class medians becomes marked. Table V*B* shows that the medians of the number of rights at the end of the first five minutes for the four forms in Grade V vary by as much as four examples. In Grade VI the range is five examples, while in Grade VII it is 2.5 examples.

Variability in Individual Performance.—Table VI shows the median differences in the individual scores between successive forms

of the test. In Grade V, for example, the median of the Arithmetical differences in scores on Forms I and II is 1.93; in Grade VI the differences in scores between Forms II and III were marked, the median being 3.13. This means that half of the children in Grade VI changed their score from Form II to Form III by three or more points. The average difference between successive forms for the total group is 2.36. This is interesting when one recalls that the average of the half year increments in norms for the test from Grades V through VII is 2.04. The changes on successive forms for half of the children equalled or exceeded .58 of the increment representing a year's change in the norms.

TABLE VI.—SHOWING MEDIAN DIFFERENCES IN NUMBER OF RIGHTS BETWEEN SUCCESSIVE FORMS

	Grade V	Grade VI	Grade VII	Total group
Form II minus Form I	1.93	2.10	2.67	2.22
Form III minus Form II	2.75	3.13	1.90	2.55
Form IV minus Form III	2.75	1.97	1.92	2.18
Form I minus Form IV	2.62	2.25	2.55	2.50
Average	2.51	2.36	2.26	2.36

When we realize that this degree of instability exists in a single measure of a pupil's ability we are reluctant to recommend that such measures be used as the only basis for promotion or classification of children. When any form of a test has an even chance of producing a score differing from that of a duplicate form, without the intervention of time or teaching, by an amount exceeding half a year's growth we hesitate to attribute to it any considerable degree of finality.

CONCLUSIONS

This study of the arithmetic variability of children seems to justify the following conclusions:

1. In four-minute addition practices using materials as reliable as single columns of nine digits the chances are even that the medians of a group for number of attempts on any day will vary from those on any other day by two examples or more, and that the medians of the

numbers of correct additions will vary from day to day by 1.3 examples or more.

2. The chances are even that in so standardized a performance as single column additions any individual will vary from day to day by 2.1 column or more in speed, and by 2.5 columns or more in number right in a four-minute practice period.

3. There is a tendency for the amount of variability in single-column addition to increase with intelligence.

4. The medians of the numbers of examples done correctly on the four forms of the Woody-McCall Mixed Fundamentals are quite constant or stable, varying not more than 1.1 in any one of the three school grades tested. The medians of the numbers of examples done correctly on these four forms at the end of the first five minutes of working time are, however, extremely unstable, varying by five examples in one class and by four in another.

5. The scores of individual pupils on the four forms of the *Mixed Fundamentals* vary considerably, the chances being even that the score of any child on any form will differ from his score on another form by 2.36 or more. It is important to remember that this amount of difference is equivalent to more than half the amount of a year's growth as stated by the norms of the test.

6. Since the factor of an active time limit appears to increase variability, it seems desirable to determine the relation between the amount of variabilit yand the length of the test or practice period.

A "WORK-LIMIT" TIMING MACHINE[1]

MICHAEL WEST

The Teachers Training College, Dacca, Bengal

A "Work-Limit" Timing Machine

In *The Journal of Educational Psychology*[1] Robinson describes an instrument for timing Work-limit Group Tests, *viz.*, tests in which the score or a part of the score of the pupil consists in the time taken to complete a certain task.

The device there described consists of 31 numbered cards, 4 inches by 5, running on a double wire. The numbers are three inches high. The first card (zero) is turned over as soon as the first paper is shown up, and that paper is marked Zero: the other cards are turned over at 5-second intervals and the subsequent papers are marked according to the numbers on the cards. A constant time (from the start to the handing in of the first paper) is added to all the scores. The cards are turned over by hand.

Stone,[2] in his Reading Test uses a very similar device to that described above. The cards are worked by hand, and he suggests the employment of a "capable boy from the upper grades" for the purpose. He also adds a constant, the nature of which is not made very clear in his Directions.

This adding of a constant would greatly increase the labor of scoring a large scale test done in several different groups or classes. In the case of practice classes, in which several work-units are done in a single period, it might be liable to produce errors. The purpose of the constant appears to be to economize the number of cards. If the times are written up on a black-board, there is no need to use a constant. The only disadvantage of the blackboard is that when the Mean Time comes along, there is rather a rush, and the blackboard operator has hardly time to write up the numbers. If however the 20 or 30 numbers round about the mean time be written up on the board before the test, the times up to and after these scores can be

---

Note.—This work was done in conjunction with my brother C. F. M. West, for whose assistance I am extremely grateful.

[1] Robinson, B. W.: A New Timing Device for Work-limit Group Tests. *Journal of Educational Psychology*, Vol. XIII, No. 8, November, 1922.

[2] The Stone Series of Narrative Reading Tests (Directions), 1922.

written up individually as required, but at the rush time the operator has only to point.

This method works very well in practice, but it suffers from the same disadvantage as the two methods noted above, namely that it requires an assistant in addition to the teacher or test-director. In the case of a test this does not matter much, for a test is usually an important and rather costly function and a not very frequent occurrence, so that a special staff can always be mobilized for it. But in the case of a practice class, which goes on daily for some considerable period, the need of an assistant is a serious objection. A method of teaching which requires two persons, a teacher and an assistant, is impracticable in any ordinary school.

For test-work it is a convenience, but for practice-work it is an absolute necessity to have an automatic system of time-recording which makes it possible for one teacher to conduct the work unassisted.

The machine required is one which will show figures at least two inches high, from 1 to 999 at regular intervals. Ideally it should be possible to vary the interval for different types of work; but, if the interval must be fixed, it should be 6 seconds in order that the Unit-scores may, by the insertion of a decimal point, be converted into minutes without arithmetical labor. The machine must be reasonably silent, and be readily set back to zero.

The turn-over type of action as used by Robinson was considered; a similar mechanism was used in a clock of German manufacture which appeared on the market about 1910: the idea was rejected as likely to prove costly and difficult to make. A Maltese-cross action running a chain of number boards was considered and rejected as not giving a sufficient range of numbers. A "Frena" camera action would give a good range but would be troublesome to zeroize. A machine consisting of three revolving dials was decided upon.

The first model constructed was driven by clock-work, with the idea of making it portable, but the machine had not power to turn number-discs of sufficient size, and was very noisy. It was therefore "scrapped."

If a weight-driven model were to be used, one might just as well, from the point of view of portability, have an electric model running from an accumulator. In the second model, therefore, the original transmission of the clock-work model was retained, but the machine was actuated by a solenoid driven from contacts on a clock. It was

found that the light contact possible on a clock, with the large current required to actuate the solenoid, was liable to "chatter" and give double strokes. Hence the clock was made to actuate a relay.

The instrument was a failure owing to its mechanical "tens"-transmission which required so much power on the double transmission (*e.g.* 99 to 100) that it over-ran on a single transmission or a simple unit stroke, and none of the checking devices used on calculating machines succeeded in preventing this. This machine was therefore also "scrapped."

In the third attempt each dial was actuated by an independent solenoid, the tens-transmission being electrical. The difficulty encountered in this machine was that the contact on the unit solenoid was so short that it would not give the ten and hundred solenoids time to act. This, however, was overcome by making the unit solenoid merely start the movement which was completed by an extra contact on the ten and hundred solenoids.

This model was the final.

The machine works silently and is capable of giving any time-unit required. As at present fitted it gives 4, 5, or 6 seconds but by a very rapid and easy adjustment it could be made to give the choice of any other three intervals preferred. The electrical "tens"-transmission has the special advantage that the dials can be very readily set to zero by turning each backwards independently, a matter of a few seconds only, whereas in a mechanical transmission zeroizing is always difficult.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY C. O. MATHEWS

Intelligence Testing

The Effect of School Training and Special Coaching on Intelligence Tests. A. E. Chapman. The Forum of Education, 1925, November, 172–183. Under improved teaching pupils intelligence scores were higher; coaching produced large gains.

Personnel Work—A New Method of Administrative Procedure. C. F. Malmberg. School and Society, 1925, Feb. 21, 236–242. Use of intelligence tests and opinions of leading educators on the advisability of informing students of their intelligence ratings.

School Progress and Mental Test Ability. Hornell Hart. School and Society, 1925, Feb. 7, 181–182. Correlation between mental-test quotients and age-grade quotients shows that stupid children are retarded but superior children are not correspondingly advanced.

Equating College Marks. Percival M. Symonds. Educational Administration and Supervision, 1925, February, 118–124. "A Scheme for equating college marking by utilizing the results of an intelligence examination to suggest to the instructor his distribution of marks."

Teachers Marks and Intelligence. Samuel P. Unzicker. Journal of Educational Research, 1925, February, 123–131. Shows the distribution of marks in the principal junior high school subjects for 425 pupils grouped according to mental ability.

Mental Ability of Freshmen High School Pupils. E. E. Keener. Journal of Educational Research, 1925, February, 113–123. Results of 12,652 pupils of Chicago high schools on the Otis Self-administering Test of Mental Ability.

Achievement Testing

Achievement in Chemistry. Stephen G. Rich. School Science and Mathematics, 1925, February, 145–149. A comparison of the achievement of high school and college students on several standardized tests.

The Need and Use of a Scientific Measure of the Results of the Teaching of Chemistry. Henry L. Gerry. School Science and Mathematics, 1925, February, 157–168. Enumeration and discussion of the values of standardized tests of achievement.

A Study of the Factors of Success in First-year Algebra. Edwin W. Schreiber. The Mathematics Teacher, 1925, February, 65–78. The first part of a study making use of standardized tests to determine factors of success. Data shows distribution of abilities.

Some True-false Examinations for Use in General Mathematics. Horace C. Wright. Mathematics Teacher, 1925, February, 83–91. Citations from others of arguments for and against the true-false test and a true-false test in general mathematics.

Hurdle Tests in Algebra. Kate Wentz. School Science and Mathematics, 1925, February, 132–144. A series of difficulty tests in first-year Algebra with criticisms.

Testing the Efficiency of the Group Method. Theodora M. Thie. The English Journal, 1925, February, 134–137. The use of the Thorndike-McCall Reading Test and Nassau County Supplement to the Hillegas Scale to compare two methods of teaching English.

The Diagnostic Value of Standardized Reading Tests. John A. O'Brien. The Catholic Educational Review, 1925, February, 65–79. Outlines four steps in the use of standardized tests and shows results of remedial teaching on silent reading test scores.

The Variability of the Accomplishment of Pupils of the Same Mental Level. T. L. Torgerson and Irene Shuman. Journal of Educational Research, 1925, February, 132–136. Tends to show that the AQ varies directly with the IQ in groups of the same mental age.

Dwight Reasoning Test. Charles A. S. Dwight. Education, 1925, February, 338–344. Various types of argument in which the student is to detect the errors and explain. Suitable for high school Seniors and college Freshmen and Sophomores.

Learning

Repetition of Errors in Algebra. Flora L. Scott. Mathematics Teacher, 1925, February, 92–96. Children tend to make similar errors over and over much more often than unlike errors.

Difficulties in Problem Solving. P. R. Stevenson. Journal of Educational Research. 1925, February, 95–103. A study of six important causes of failure in arithmetic with suggested remedial measures.

Fatigue Tests and Incentives. Helen Francis Whiting and Horace Bidwell English. Journal of Experimental Psychology, 1925, February, 33–49. Fatigue tests show no significant decrease in efficiency. Hypothesis advanced that fatigue is primarily an emotional appetite while exhaustion affects the mechanism.

Relative Performance of College Students as Conditioned by Time of Day and Day of Week. Donald A. Laird. Journal of Experimental Psychology, 1925, February, 50–63. Shows variations in the work curve for days of the week and hours of the day in speed, comprehension and retention of reading matter.

Character and Personality

Tests of Personality Traits. Frank N. Freeman. The School Review, 1925, February, 95–106. A fourfold classification of personality traits and some results of Will-temperament Tests.

Ideals, Situations, and Trait Actions. W. W. Charters. The Elementary School Journal, 1925, February, 425–436. A classification of situations involving honesty.

MISCELLANEOUS

The Psychology of the Fine Arts. Charles H. Judd. The Elementary School Journal, 1925, February, 414–423. The last of a series of articles on educational psychology analyzing social institutions. The evolution and importance of music is used as an example of the fine arts.

Reading Interests Compared with the Content of School Readers. Emma B. Grant and Margaret L. White. Teachers College Record, 1925, February, 480–497. A study of the content of fifteen school readers for primary grades.

Notes on the Interpretation of Correlation Measures. E. R. Hamilton, 1925, November, 193–199. Some of the dangers in the collection and interpretation of statistical data are pointed out.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY JOHN HOCKETT

The Lincoln School of Teachers College

An Introduction to Child Psychology

The Growth of the Mind, by Kurt Koffka. Translated by Robert M. Ogden. New York: Harcourt, Brace & Co. 1924. Pp. XVI + 382.

The purpose of the book is definitely stated by the author: "The general aim of my book has been but to point out a way in which a solution of these numerous problems (behavior and development) can be attained. The nature of mental development as it has been revealed to us is not the bringing together of separate elements, but the arousal and perfection of more and more complicated configurations in which both the phenomena of consciousness and the functions of the organism go hand in hand." This seems to be the best single statement of the hypothesis of the *Gestalt Psychologie* of which the present work is the most complete statement in our language.

Lack of space prohibits reference to, and comment upon, many problems and their interpretations that are presented. In addition to the discussion of the general problems of mental life, methods of study, starting point, development and maturation, and the mind of the child, special interest will be found by many in some of the briefer sections. These include a resumé of the work of the behaviorists; an interpretation and criticism of Thorndike's investigations and of his laws of learning; a summary of Ruger's studies of learning three dimensional puzzles; and a description of Köhler's recent experiments with the learning of chimpanzees. Throughout, the hypothesis of configuration, as opposed to the atomistic, is advanced as the best explanation of the growth of the mind. The point of view is somewhat new to the American student; though in fact it might be difficult to find any recognized psychologist who will admit holding an atomistic viewpoint. The fact that analysis and synthesis are easy methods of

description is probably a sufficient explanation to account for this trend in psychological literature. It is believed, though, that the work will prove to be a real contribution, not in the sense that it overthrows the body of psychological principles now generally accepted, but in that it gives new emphasis to a phase of psychology which, if not so much in thought, perhaps more in writing, has been relatively underemphasized. In this it may be likened somewhat to the contributions made by Watson. However, it seems doubtful whether the Gestalt psychology will have any great following. We tend to be pledged to objective and analytic methods and prefer them to the somewhat more philosophical viewpoint in the present work.

" . . . The Author's skill in executing his task, and his mastery of the results thus far achieved by experimental work in the field of child psychology, have been so happily combined as to insure a lively interest in his book. It is therefore gratifying to me, personally, to have been instrumental in making the book available to English and American students of Educational Psychology" says Robert Morris Ogden, the translator of the work. It is with somewhat a similar feeling that the reviewer enthusiastically calls attention to it as one of the very interesting works in psychology which have appeared recently.

EDWIN MAURICE BAILOR.

THE EDUCATION OF THE MENTALLY INFERIOR

The Education of Handicapped Children, by J. E. Wallace Wallin. Boston: Houghton Mifflin Co., 1924. Pp. 394.

With the increasing geographical spread of the recognized need for differential education has come a desire on the part of educators to be informed of the progress along specific lines. They wish for a general survey in each field of specialized education, of the initial problems, and also the general prospects. Dr. Wallin has met this need in respect to the education of children who are mentally inferior.

The material is presented in three parts. Part I gives the "History of the education of handicapped children," and "The psychological theories and definitions of feeble-mindedness and backwardness." It is historical and explanatory. It includes a discussion of constructive work for the mentally handicapped which is a matter of such recent development that it treats primarily of the present.

Part II deals with "Public School questions relating to the education of handicapped children." It covers practically every large problem from organizing special instruction for such classes (including details of the curriculum), administering and supervising them, to "The qualifications needed by the teacher of the mentally deficient." The material is from a very practical point of view. It contains much that is concrete. Many suggestions are offered. They impress one forcibly as emanating from the experiences of the actual participant in the education of handicapped children. They present various approaches to the problems, from the view point of the administrator, teacher, visiting teacher, social worker, psychologist, and other educationists. They are so numerous that practically all phases of the work are touched.

Part III is concerned with the social aspects of the problem, "The social menace of the feeble-minded," and "The program of constructive endeavor in their behalf." Facts are discussed concerning the relation of feeble-mindedness to (1) high fecundity, (2) degenerate progeny, (3) criminality, (4) inebriety, (5) social immorality, and (6) pauperism and vagrancy. The part is concluded with a discussion of "Prevention and elimination rather than cure," and "The proper and humane support, control, and training of the feeble-minded already in existence."

An appendix gives technical matter on the classification of mentally deficient children, and a list of the complete titles of writings referred to in the text.

The real contribution is in Part II which presents so much material from the writer's professional experience. It is not theoretical. It outlines the various problems as they actually exist. It indicates how numerous and serious they are, and suggests ways of meeting them. It rings with a note of assurance to all those who are about to enter the work with children handicapped through mental inferiority. It stimulates a broader viewpoint in those who are already at work in the field. Hasty reading of this part as a whole gives one the scope of this type of differential education. Careful reading leaves one with a good plan for studying or with many good suggestions for meeting the problems in actuality.

The book is one which should be on the teachers' reading shelves in school systems contemplating the organization of special classes for the handicapped, as well as on shelves for general reading of others interested and definitely concerned in the general problem. For the

initiated or others who are studying the problem of mental deficiency or allied conditions the material is suggestive in the way of a general outline of the topic. GRACE A. TAYLOR.

A Tentative Inventory of the Habits of Children from Two to Four Years of Age, by Andrus Ruth. New York: Teachers College, Columbia University, Contributions to Education, No. 160, 1924. Pp. 50.

This is a report of an interesting exploration into the study of the preschool child by the inventory method. An inventory containing 447 items distributed under four headings—emotional, mental, motor, and social-moral—was used. The purposes of the study were: (1) To discover what abilities children of nursery school age could be expected to possess, and (2) to evaluate the inventory method as a means of training teachers in more accurate observation of children of these ages.

Diaries of 52 children in four nursery schools in New York City were kept by 69 students from a class in kindergarten-primary education, each student observing one child for 15 hours. The diary results were then transferred to the inventory. The experimenter also filled out inventories from these diaries and checked the students' scores against her own. She found that training was necessary to accurate results, but that with training reliable scores could be obtained from such a class with three hours a week of observation.

Intercorrelations between the four divisions of the inventory are given for the two-year groups separately. The correlation between the "mental" division and the Kuhlmann Intelligence Test was .661 for age two and .611 for age three.

Individual profiles are given for some selected children (so-called neurotic, high IQ and low IQ) and group profiles are also given.

The final revised inventory contains 1230 items distributed under the four mentioned headings. This report is a distinct contribution to the field of child study at the preschool age. BETH WELLMAN.

High School Experimentation

Studies in Secondary Education, II, by the Faculty of the University High School of the University of Chicago. Chicago: The University of Chicago, 1925. Pp. 202.

This second monograph on the Chicago University High School carries forward and further particularizes the account of what has been going on there since Dr. Henry C. Morrison became Superintendent of the Laboratory Schools in 1919. There are 10 reports this time, contributed by 12 members of the faculties of the high school and the school of education. The first six deal with pupil-progress accounting, and the last four, with experiments in curriculum organization and administration. In Part I such matters as the following are presented: The evolution of a marking and commendatory report system which sets a premium, not on grades or points or credits, but on mastery in learning minimal essentials and upon intellectual interest and ambition. Provisions for enriching the experience and hastening the progress of superior students, and measures of the success of these provisions. Progress of experimental and control groups through the practice tests of a unit of elementary physical science, including some very significant data on the relation of intelligence to progress, the fluctuations in progress from week to week, and the identification of factors in progress. Data derived from specially devised and from standard tests as to the progress of pupils in the Community Life English course, showing the effect of the course on reading rate and assimilation, extensiveness of outside reading, and formal accuracy and general merit in composition. The development of appreciation in upper-class English through the voluntary project, with striking examples of creative work in epic, libretto, and lyric. The effect of specific practice exercises in eliminating errors in English usage as shown in papers written for English and for science, including the account of a laborious but excellent technique for measuring this effect. The construction of scientifically and statistically defensible tests for the measurement of achievement in French-aural comprehension, vocabulary, grammar, and translation.

Part II contributes to the curriculum field such accounts as the following: The evolution of advanced courses in social science for high school seniors through successive trials of materials organized around current events, business and society, and modern problems, to the present tentative formulation of two semester courses in the

elements of political science and political economy respectively, in which fundamental principles and orderly habits of thinking are emphasized instead of the acquisition of ephemeral and expansive information. A special course on *The Worker in Modern Society*, suggestive for the enrichment of the curriculum in high schools where time may be found or where sufficient pupils may be free to elect. A restatement of Mr. Breslich's ideas about first courses in secondary school mathematics. A decidedly fruitful report of the organization and administration of a first year French course in the Junior College.

There is valuable material in this volume for the classroom teacher, the supervisor, the counselor, the measurement expert, the curriculum maker, and the administrator. As in the first volume of these reports, the discussions are restrained, clear, and concretely illustrated. One receives again the strong impression of a well integrated institution dominated by a clear-cut philosophy of learning, which each year finds more specific, more confident, and more pervading expression in both administration and teaching. One may quarrel with this philosophy and its application if he wish, but he cannot but regard very highly such intelligent and conscientious experimentation with it as the reports of this monograph disclose. M. H. WILLING.

PROGRESSIVE THEORY AND PRACTICE IN BELGIUM

The Decroly Class, a Contribution to Elementary Education. By Amelie Hamaïde. Translated by Jean Lee Hunt. E. P. Dutton & Co., 1924. Pp. XXVIII + 318.

Most philosophers who have dealt with reorganizing the materials of a course of study, or with limiting them, or adding to them, have met with opposition on the part of the public schools. They have had great difficulty in getting any of their ideas carried out under anything like regular public school conditions. Mlle. Hamaïde has, therefore, made a real contribution when she shows how she was able to develop in a public school system the philosophical principles which underlie the thought of Ovide Decroly. The conditions under which she herself worked were unusually favorable but she was able to supervise ten other classes in the public schools of Brussels where conditions were quite normal. The exposition which she gives of her experiment forms the material for *The Decroly Class*.

Ovide Decroly in Belgium, just as John Dewey, in America, recognizes adequately the nature of the child and the way in which that nature develops. He believes that activities which are real and full of suggestive meaning for comprehending the fundamental needs of life are essential for such growth. Decroly requires on the part of children observation of the various phases of life activities as an appropriate means for developing and clarifying them. Observation with him must be pointed, it must serve a purpose in the actual development of a center of interest, as Mlle. Hemaïde calls his activity unit. After observation comes association of the varied elements which have been considered during the course of the development of the activity. It is through this association that the pupil classifies his information and puts it in its proper relation. Decroly provides for two kinds of expression, concrete expression through such means as modeling, painting, or constructing, and abstract expression through reading, writing, and numbering. All these Mlle. Hamaïde carefully explains in the first part of her book.

Type programs of centers of interest are suggested for grades one, two and three. Outlines are prepared for the development of each of these and it is shown how the so-called formal subjects of study grow out of them. Perhaps the most questionable feature in the book is the fact that the selection and placement of these centers of interest are merely matters of opinion.

The most valuable parts of Mlle. Hamaïde's book are the reports of individual children and the specimens of work. The reports of the individual children show that growth in every case was most remarkable and that the accomplishment of social values was no less well attained than was the mastery of the tool subjects. On the whole the objective evidence which Mlle. Hamaïde presents offers great encouragement to those in America who are working in progressive schools. Her report is a most valuable contribution not because it presents a new point of view to progressive thinkers in America but because it shows how a valuable new philosophy can be taken over and carefully developed under fairly normal school conditions. In the latter particular it ranks with such books as Meriam's *Child Life and the Curriculum*, Colling's *An Experiment with a Project Curriculum* and Pratt and Wright's *Experimental Practice in the City and Country School.*

JAMES S. TIPPETT.

Experiments in Learning

Nature and Development of Learning Capacity, by William H. Pyle. Baltimore: Warwick and York, 1925. Pp. 122.

This monograph very naturally follows Professor Pyle's *Laboratory Manual in the Psychology of Learning* (Warwick and York, 1923). It is a report of a series of learning studies which were directed by the author and financed by the University of Missouri. The tests used apparently were selected from those presented in the Laboratory Manual and include, among others, the now familiar experiments of card- and marble-sorting, a substitution test, mirror drawing, and paragraph comprehension. The particular problems upon which attack was made include those of the general or specific nature of the learning functions; sex differences in learning; sex differences in variability in learning; the nature of the differences in learning capacity between good and poor learners; and racial and other differences in capacity to learn. The monograph closes with a brief chapter which is a reprint of an article which appeared in the *Psychological Review* for July 1924 on "A Theory of Learning" which is the usual one of connections. Three laws of brain action are suggested as adequate to "account for all the facts in the case of bonds of simultaneous experience: (1) Simultaneous brain activities are connected; (2) only in focalized brain activities are strong connections formed; and (3) the more intense and vigorous the focalized brain action, the stronger are the associative bonds that are formed."

Much work and considerable expense must have been involved in the study. Contributions of data are always interesting and valuable. However little is presented that seems new. The problems chosen are of great importance; but the tests, particularly in the absence of controls and adequate treatment of data seem to assist little in their solution. The monograph deserves commendation, but the reviewer felt a pang of disappointment when he examined the selection and treatment of the experiments after having been attracted by the ambitious title "Nature and Development of Learning Capacity."

Edwin Maurice Bailor.